Spillover Effects of Early-Life Medical Interventions*

N. Meltem Daysal
University of Southern Denmark and IZA

Marianne Simonsen
Aarhus University and IZA

Mircea Trandafir
University of Southern Denmark and IZA

Sanni Breining
Ramboll Management Consulting

February 2019

Abstract:

We investigate the spillover effects of early-life medical treatments on the siblings of treated children. We use a regression discontinuity design that exploits changes in medical treatments across the very low birth weight (VLBW) cutoff. Using administrative data from Denmark, we find that siblings of focal children who were slightly below the VLBW cutoff have higher 9th grade language and math test scores. Our results suggest that improved interactions within the family and parental compensating behavior may be an important pathway behind the observed spillover effects.

Keywords: Medical care, birth, children, schooling, spillovers

JEL Classifications: I11, I12, I18, I21, J13

* Daysal: University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: meltem.daysal@sam.sdu.dk); Simonsen: Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark (email: msimonsen@econ.au.dk); Trandafir: University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark (email: mircea.trandafir@sam.sdu.dk), Breining: Ramboll Management Consulting, Denmark (email: sannibreining@gmail.com). Doug Almond, Prashant Bharadwaj, Aimee Chin, Gordon Dahl, Nabanita Datta Gupta, Joe Doyle, Mark Duggan, Bill Evans, David Figlio, Kristiina Huttunen, Katrine Løken, Bhash Mazumder and seminar participants at Bergen, Concordia, Gothenburg, Houston, IZA, Mainz, Michigan, Tilburg, VIVE, York, Zurich, 2nd SDU Workshop on Applied Microeconomics, SFI-Lund Workshop on Health Economics, Essen Health Conference, and Copenhagen Education Network provided helpful comments and discussions. Breining and Simonsen gratefully acknowledge financial support from CIRRAU. The authors bear sole responsibility for the content of this paper.

1. Introduction

It is well understood in economics that the family plays a central role in the human capital accumulation of children (Cunha and Heckman, 2007; Cunha et al., 2010; Almond and Currie, 2011). While the majority of the economic literature focuses on the effects of parent-child interactions, a growing number of studies now emphasize the importance of sibling interactions in shaping academic achievement. The evidence suggests that older siblings' education choices and test scores causally affect younger siblings' academic outcomes (Nicoletti and Rabe, forthcoming; Joensen and Nielsen, 2018). The existing research also indicates that children's early-life health endowments impact the academic outcomes of other children in the family by changing parental investments. Evidence of such spillover effects are found both in developing and in developed countries and the magnitudes of the effects are economically large (Yi et al., 2015; Black et al. 2016). A natural question then is whether – and by how much – interventions that improve child health affect these sibling spillovers on child human capital. In this paper, we address this understudied question by investigating the spillover effects of early-life medical treatments on the human capital accumulation of siblings, focusing on the specific case of treatments provided to very low birth weight (VLBW) children, i.e., children with birth weight below 1,500 grams.[1]

Medical interventions targeting VLBW children constitute an ideal setting to study spillover effects for several reasons. To begin with, they have been found to substantially improve health (e.g., Cutler and Meara, 1998; Almond et al., 2010; Bharadwaj et al., 2013) as well as academic achievement (e.g., Bharadwaj et al., 2013). Second, although VLBW children represent a small share of all births, they account for a substantial portion of newborn health care expenditures. For example, VLBW babies in the US represent around 1.5% of all births but only the neonatal intensive care unit costs associated with these babies account for 30% of newborn health care costs (Johnson et al., 2013). Finally, focusing on treatments provided to VLBW children allows us to overcome identification challenges arising from potentially correlated unobservables within the

[1] The economic literature on returns to early-life medical interventions almost exclusively studies the effects on treated children (e.g., Cutler and Meara, 1998; Chay et al., 2009; Field et al., 2009; Almond et al., 2010; Bharadwaj et al., 2013; Daysal et al., 2015; Hjort et al., 2017; Bütikofer et al., forthcoming; Daysal et al., forthcoming). We are aware of only one study on spillover effects with a causal interpretation. Adhvaryu and Nyshadham (2016) examine the effects of a large-scale iodine supplementation program in Tanzania and find that the siblings of treated children were more likely to be immunized.

family (e.g, shared genetic factors that impact both sibling outcomes and the receipt of medical treatments by targeted children). Specifically, we use a regression discontinuity design that exploits changes in medical treatments across the very low birth weight threshold to address the non-random assignment of medical treatments (Almond et al., 2010; Bharadwaj et al., 2013). We investigate separately siblings of focal children (i.e., children with birth weight in a small window around 1,500 grams) with gestational ages above and below 32 weeks because the medical guidelines prescribe additional medical treatments to children with gestational age below 32 weeks regardless of their birth weight.[2]

Using population-level data from Denmark, we find substantial positive spillovers on the test scores of siblings of focal children with gestational age of at least 32 weeks. Our estimates suggest that siblings of focal children who were slightly below the VLBW cutoff have higher 9th grade language and math test scores by 0.386 and 0.255 standard deviations, respectively. These results are economically large. The effects, in fact, correspond to a reduction of 33-69% in the test score gap between children in the top and bottom 5% of the income distribution. However, the improved academic achievement of siblings does not translate into a higher probability of enrollment in education beyond the 9 years of compulsory schooling, likely due to the already high rate of high school enrollment in the sample (78%).[3]

This empirical design as well as the richness of the Danish register data allow us to conduct a range of robustness and falsification checks. For example, we find no evidence of discontinuities at the VLBW cutoff in the test scores of siblings of children with a gestational age of less than 32 weeks. Similarly, we show that there are no discontinuities in the early-life health outcomes (28-day and 1-year mortality) of older siblings, indicating that unobserved family traits are unlikely to drive our results.

In the last section of the paper, we turn to the potential mechanisms behind the observed spillover effects. We first provide evidence suggesting that siblings' direct exposure to treatments or general improvements in parental health education are unlikely to be the main driver of our results: we find no discontinuities in health outcomes (hospitalizations and ER visits) of siblings. We next replicate the findings in the previous literature that focal children who weigh slightly less than

[2] Here and in the rest of the paper we refer to children with birth weight around 1,500 grams as "focal children."

[3] During our study period, Denmark had nine years of compulsory education. Loosely speaking, high school included grades 10-12.

1,500 grams are more likely to survive within the first year of life and to have better test scores in 9th grade. Our estimates are remarkably comparable in magnitude to the reductions in infant mortality and improvements in test scores from previous studies (Almond et al., 2010; Bharadwaj et al., 2013). We also find that the medical treatments provided to VLBW children do not affect the likelihood of common childhood disabilities (e.g., ADHD, intellectual disability, behavioral and emotional disorders, epilepsy, cerebral palsy), but they seem to result in better health during the school years, as proxied by reduced hospital and ER visits.

In order to understand how the improvements in focal child health and academic achievement might impact siblings, we check whether there are any changes in total household resources or in several indicators of the quality of the family environment. We do not observe discontinuous changes across the VLBW cutoff in total household resources, including personal and total income, and labor force participation. Our results instead suggest that early-life medical interventions may improve the general family environment. In particular, we find that the mothers of VLBW focal children have better mental health soon after the focal child is born. In addition, we find evidence of heterogeneity in the spillover effects on sibling academic achievement by sibship characteristics that are most closely tied to the quality of peer interactions (gender of sibling, gender composition). Finally, we also provide indirect evidence on the role of parental reallocation of resources within the family. If the production of human capital exhibits dynamic complementarities, then parental investments have higher returns for children with high initial endowment than for children with low initial endowment. If parents engage in compensating behavior, then our results suggest that the siblings of VLBW focal children should receive more parental resources than the siblings of focal children with birth weight slightly above 1,500 grams. In our data, we find that siblings of VLBW focal children experience higher test score gains if they have high initial endowments themselves (as proxied by their birth weight), a pattern consistent with compensating behavior by their parents.

Our study makes three contributions. First, it adds to our understanding of the determinants of human capital accumulation. The economic literature generally considers the effects of various interventions on the educational outcomes of targeted children. Our findings suggest that human capital accumulation may also be affected by interventions targeted at other family members. Second, we contribute to the discussion on the cost effectiveness of early-life medical treatments. During the past few decades, medical spending for the very young increased substantially faster than spending for the average individual. For example, US annual spending on individuals aged 1 to 64 increased by 4.7 percent between 1960-1990, while per capita spending on infants under 1 year old increased by 9.8 percent per year (Cutler and Meara, 1998). Technological innovations are widely considered the main driver of this medical cost growth, both in general and in the specific case of early-life treatments (Newhouse, 1992; Cutler and Meara, 1998). As medical expenditures keep increasing, understanding the efficacy of early-life medical interventions becomes even more important. Our results indicate that medical treatments for VLBW children have externalities to other family members that raise their net benefits. Finally, our results speak to the economic literature that relies on sibling fixed effects models to account for unobserved heterogeneity across households. To the extent that siblings have positive spillovers on each other, sibling fixed effects models would underestimate the true treatment effects.

2. Institutional Background

The majority of Danish health care services, including birth related procedures, are free of charge and all residents have equal access (Health Care in Denmark, 2008). The first European neonatal intensive care unit was established in 1965 at Rigshospitalet in Denmark and the use of early-life medical technologies has since followed the international development (Mathiasen et al., 2008).

Danish neonatal medicine textbooks pay particular attention to VLBW children (i.e., children weighing less than 1,500 grams, regardless of gestational age) and very premature newborns (i.e., those with a gestational age less than 32 weeks, regardless of birth weight). These birth weight and gestational age classifications are frequently found in medical research papers based on Danish data where the focus is often on their higher mortality rates (e.g., Thomsen et al., 1991; Hertz et al., 1994). Medical handbooks suggest courses of treatment based on either birth weight or gestational age (Schiøtz and Skovby, 2001). Specific recommendations in terms of nutrition and vitamin supplements exist for VLBW children (Peitersen and Arrøe, 1991). In addition, papers indicate that children below 1,500 grams or born before 32 weeks of gestation are more likely to receive additional treatments such as cranial ultrasound (Greisen et al., 1986), antibiotics (Topp et al., 2001), prophylactic treatment with nasal continuous positive airway pressure (nCPAP), prophylactic surfactant treatment and high priority of breast feeding, and use of the kangaroo method (Jacobsen et al., 1993; Verder et al., 1994; Verder, 2007; Mathiasen et al., 2008).

Anecdotal evidence from hospital and regional specific notes also outline special services that are provided to families with children below 1,500 grams or below 32 weeks of gestational age. These services include referrals to a physiotherapist who guides and instructs parents on how to stimulate the development of the child and on various baby exercises. It is also mentioned that all children below 1,500 grams or below 32 weeks of gestational age are routinely checked 1-2 months after discharge and again when they are five months, one year and two years old.[4]

3. Conceptual Framework

Early-life medical interventions provided to VLBW children may influence human capital accumulation of their siblings both directly and indirectly. As discussed in the previous section, VLBW children benefit from additional medical resources. These resources could directly improve the health (and thus academic achievement) of siblings if they are also exposed to the treatments (e.g., increased routine checks) or if the treatments help parents understand the role of different health inputs.

Siblings may be impacted indirectly through changes in VLBW child outcomes too. Medical interventions early in life have been shown to improve the survival, short-term health and later-life academic achievement of treated children.

The improved focal child health may impact sibling academic achievement because previous literature links child health to resources available within the family. For example, parents of children in better health tend to work more (Powers, 2003; Corman et al., 2005; Noonen et al., 2005; Wasi et al., 2012; Kvist et al., 2013). While this may increase total family income, it might as well reduce available time for parent-child interactions both for the focal child and for their siblings.

In addition, focal child health may lead to changes in intra-household resource allocation that impact sibling educational outcomes. A large literature in economics documents that parental investments are a function of children's early life endowments (see Almond and Currie, 2011, Almond and Mazumder, 2013, and Almond et al., forthcoming, for a review of this literature). Empirical evidence on how parents change their resource allocation is mixed. Some studies find that parents tend to reinforce differences in early life endowments (e.g., Rosenzweig and Wolpin, 1988; Behrman et al., 1994; Parman, 2015) while others find evidence of compensating behavior (Behrman et al., 1982; Pitt et al., 1990; Bharadwaj et al., forthcoming; Yi et al., 2015; Adhvaryu and Nyshadham, 2016).

[4] Unfortunately, our data does not include any information on specific early-life treatments.

Previous literature also finds an association between child health and changes in family environment. For example, poor child health is linked to higher likelihood of family dissolution (e.g., Corman and Kaestner, 1992; Reichman et al., 2004; Kvist et al., 2013), which is in turn tied to worse outcomes for all children in the family (e.g., Manski et al., 1992; Haveman and Wolfe, 1995; Ginther and Pollak, 2004). Similarly, child health is associated with parental well-being. The extant literature shows a positive association between child mortality and the risk of psychiatric and physical health problems of parents (e.g., Levav et al. 2000; Li et al., 2003; Li et al., 2005), which are important inputs in the development of all the children in the household. Thus, improved focal child health may lead to better sibling academic outcomes through an improved family environment.

Finally, sibling outcomes may be impacted through changes in the quality of peer interactions. Previous psychological studies suggest that older children may act as role models for younger siblings (e.g., Dunn, 2007). This is consistent with the economic research linking younger siblings' educational outcomes and risky behavior to their older siblings (e.g., Oettinger, 2000; Ouyang, 2005; Altonji et al., 2016) and suggests that health and academic achievement gains resulting from early-life medical interventions may have positive spillovers on siblings, especially the younger.

Overall, this discussion indicates that the direction of the spillover effects of early-life medical interventions is theoretically ambiguous and ultimately an empirical question.

4. Empirical Strategy

The goal of this paper is to estimate the effect of early-life health interventions on the human capital accumulation of siblings of targeted children. Identification of these effects is complicated by the non-random assignment of medical treatments. In particular, there may be unobserved determinants of sibling outcomes that are correlated with the receipt of medical treatments by targeted children, such as shared genetic factors. In order to address this endogeneity, we follow Almond et al. (2010) and Bharadwaj et al. (2013) and use a regression discontinuity design that exploits changes in medical treatments across the VLBW threshold. Specifically, we estimate local-linear regressions of the form:

$$y_{ijt} = f\left(bw_j - 1500\right) + \beta VLBW_j + \delta X_{ijt} + \epsilon_{ijt} \tag{1}$$

where y_{ijt} is an outcome of sibling i of focal child j at time t after the birth of the focal child, bw_j is the birth weight of focal child j, $f(\cdot)$ is a first-degree polynomial in our running variable (distance to the VLBW cutoff) that is allowed to differ on both sides of the cutoff, $VLBW_j$ is an indicator for focal child j having very low birth weight (i.e., $bw_j < 1500$), and X_{ijt} is a vector of covariates.[5] The parameter of interest, β, is an intention-to-treat estimate of the effects that additional medical treatments received by VLBW newborns may have on their siblings.[6]

As discussed in Section 2, newborns with a gestational age of less than 32 weeks are always covered by the medical guidelines for receiving additional medical interventions, irrespective of their VLBW classification. Since there is no discontinuity in the medical treatments potentially provided to these focal children (Bharadwaj et al., 2013), we do not expect to observe a discontinuity in the outcomes of their siblings either. Therefore, from here on we focus exclusively on the siblings of children with gestational age of at least 32 weeks and we use the siblings of children with gestational age below 32 weeks only as a falsification check (hereafter GA32+ and GA32-).

Our baseline regressions use a triangular kernel that assigns decreasing weights to observations farther away from the cutoff. We choose our bandwidth based on a rule-of-thumb procedure suggested by Calonico et al. (2014), which yields optimal bandwidths between 149 grams and 303 grams with an average of 225 grams.[7] We choose 200 grams as our preferred bandwidth to ensure that newborns on either side of the VLBW cutoff are nearly identical. This bandwidth is larger than the one used by Almond et al. (2010) for US data, but is the same as the bandwidth used by Bharadwaj et al. (2013) for Norwegian data and reflects the smaller number of observations available in Denmark and Norway. The vector of covariates, X_{ijt}, includes indicators for heaping

[5] Since eligibility for treatments depends on birth weight or gestational age, an alternative strategy would rely on the 32-week cutoff for gestational age. However, gestational age is recorded in full weeks in our data, making it too coarse to implement this strategy.

[6] In the Appendix we also replicate the findings in the previous literature investigating the impact of medical technologies on focal children themselves using a similar strategy to equation (1):

$$y_{jt} = f(bw_j - 1500) + \alpha VLBW_j + u_{jt}$$

where y_{jt} is an outcome of focal child j at time t.

[7] The implied optimal bandwidths are as follows: 303 (language test score), 251 (math test score), 149 (academic high school track), 167 (vocational high school track), and 258 (enrollment beyond compulsory schooling).

at multiples of 50 grams in all specifications unless mentioned otherwise (Barreca et al., 2011).[8] We conduct inference by constructing robust confidence intervals following Calonico et al. (2014, 2018). These confidence intervals are centered on bias-corrected estimates instead of the usual (conventional) estimates and use standard errors from a specification with a higher-order (in our case, a second-degree) polynomial in the running variable. Therefore, in addition to coefficient estimates and their robust standard errors, we also report bias-corrected estimates and we indicate the significance of the estimated coefficients based on these robust confidence intervals. Finally, some of our robustness checks additionally control for child and family characteristics (see Section 5 below).

5. Data

Our key data set is the Birth Register, which includes information about the universe of births in Denmark starting from 1970. For each child, the data includes information on the exact date of birth, gender, and plurality. Birth weight is recorded in 250-gram intervals between 1973-1978, in 10-gram intervals in the period 1979-1990, and at the gram level since 1991. Gestational age is added beginning in 1982. Using parental identifiers, we are able to link children to their parents and siblings and determine parity. We also link this data to other register data that provide information on both parents and children regarding demographic characteristics, labor market outcomes, health outcomes and academic achievement.

Our main outcome variables relate to human capital accumulation. We use course-specific test scores from 9th grade qualifying exams in both reading and math, available between 2001 and 2010. All exams are graded by the teacher and by an external examiner, with the evaluation of the external examiner overruling that of the teacher. To be able to compare test scores across cohorts, we standardize them to have zero mean and unit standard deviation within each cohort. We also study effects on enrollment beyond compulsory education by age 19 and on enrollment in an academic or vocational high school track.[9]

[8] Given that birth weight is measured in grams, heaping is generally symmetric around our cutoff point and hence our strategy is less likely to be affected by the criticism raised by Barreca et al. (2011). Indeed, we show in Section 6.3 that our results are robust to the exclusion of these controls.

[9] During our study period, Denmark had nine years of compulsory education. As such, enrollment beyond compulsory education is analogous to not being a high school dropout in the US.

We use data on focal children and parents when examining potential mechanisms behind the observed spillover effects. In particular, we investigate whether early-life medical interventions impact focal child physical health (28-day and 1-year mortality), mental health (diagnosis of intellectual disability or attention deficit hyperactivity disorder) and academic achievement. Similarly, we study whether treatments provided to focal children early in life impact parental mental health (antidepressant use) and labor market outcomes (income, employment, and number of days worked).[10]

Finally, some of our robustness checks control for focal child characteristics (gender, gestational age, parity, plurality, birth year, birth region), maternal characteristics at the birth of the focal child (age, years of education, marital status, immigrant status), and sibling characteristics (gender, parity, plurality, birth weight, and birth year).[11]

We define the analysis sample in several steps.[12] First, we select focal children born between 1982 and 1993.[13] We then exclude observations for which either birth weight or gestational age are missing and restrict the sample to those with birth weight within 1,300-1,700 grams. Within this sample, 3,677 focal children have siblings (hereafter the *FC sample*). We consider the siblings of these children, defined as children born to the same mother from different pregnancies. We include both older and younger siblings because the receipt of additional medical treatments around the VLBW cutoff does not seem to impact future fertility decisions.[14] We focus on siblings who are

[10] We have access to prescription drug data beginning from 1995 so we are unable to construct measures of antidepressant use for the first two years after the birth of any focal child in our sample.

[11] Maternal education is missing for a small number of observations (315 observations corresponding to 154 mothers). We replace these with the median years of education by birth cohort and include an indicator for imputed maternal education.

[12] Appendix Table A1 details the construction of our analysis sample.

[13] We restrict our sample of focal children to cohorts born after 1982, when both birth weight and gestational age are recorded in the data. We include cohorts born up to and including 1993 for two reasons. First, this allows us to have access to human capital accumulation information for all cohorts, which makes it possible to compare the effects of early-life health interventions on focal children in our context to those in previous studies. Second, evidence suggests that medical guidelines around the VLBW cutoff are less likely to be binding in recent years (see, for example, footnote 20 in Bharadwaj et al., 2013).

[14] It is possible that a focal child has more than one sibling. Our baseline regressions treat each sibling-focal child pair as an independent observation. This is not a concern for our identification because parity of the focal child and total family size are relatively smooth across the cutoff in the FC sample. In addition, we find no evidence of a discontinuity

old enough for us to observe their academic outcomes. Tests are administered when children are around 15-16 years old, so data on test scores are available for cohorts of siblings born between 1986-1997. Enrollment outcomes are measured at age 19 and include siblings born between 1970-1993. The combined sample includes 5,827 observations, of which 2,516 are siblings of focal children with gestational age of less than 32 weeks and 3,311 are siblings of focal children with gestational age of at least 32 weeks (the *sibling sample*).[15]

6. Results

6.1. Tests of the Validity of the Regression Discontinuity Design

The validity of an RD design rests on the assumption that individuals do not have precise control over the assignment variable. Since women cannot precisely predict the birth weight of their children, the variation in birth weight near the VLBW cutoff is plausibly as good as random (Almond et al., 2010; Bharadwaj et al., 2013). However, the key identification assumption of the RD design could be violated if physicians systematically misreport birth weight, especially in the presence of financial incentives for manipulation (Shigeoka and Fushimi, 2014; Jürges and Köberlein, 2015).

In order to test this assumption, we examine the frequency of births by birth weight within our bandwidth around the cutoff. Figure 1 plots the distribution of observations in the sibling sample by birth weight of the focal child, separately for siblings of GA32+ and GA32- focal children.[16] We use 10-gram bins because birth weight is reported in 10-gram intervals for most of our sample

at the cutoff when we examine the probability of having a younger sibling, the number of younger siblings, and the birth spacing between focal children and younger siblings (see Table 1 and Appendix Table A2).

[15] There are 3,324 siblings born between 1986-1997. We have data on language test scores for 2,641 siblings (1,130 for GA32- and 1,511 for GA32+) and on math test scores for 2,656 siblings (1,139 for GA32- and 1,517 for GA32+), implying that test scores are missing for approximately 21% of the eligible cohorts in the sibling sample. This is because children can be exempt from taking the test if, for example, they have a documented disability. This could be a concern if medical treatments provided to focal children impact test-taking ability of siblings. However, we find that both the probability of siblings taking the test and the age when the test is taken are smooth across the VLBW cutoff (see Appendix Table A2). We have enrollment information for all eligible cohorts, including 4,879 siblings (2,120 for GA32- and 2,759 for GA32+).

[16] Appendix Figure A1 provides the distributions of births in the FC sample for GA32+ and GA32- children, respectively.

period. Similar to previous studies (Almond et al., 2010; Bharadwaj et al., 2013), we observe reporting heaps at multiples of 50 and 100 grams but there is no evidence of irregular heaping around the VLBW cutoff in any of the samples. We check this more formally by estimating a local-linear regression similar to our baseline model, using the number of births in each birth weight bin as the dependent variable (McCrary, 2008; Almond et al., 2010). We do not find any evidence of a discontinuity in the frequency of births at the VLBW cutoff.[17] These results suggest that birth weight is unlikely to be manipulated in our context.

In the remainder of this section, we check whether there are differences in observable characteristics across the VLBW cutoff by estimating our baseline model with the covariates as dependent variables. If the RD design is valid, then there should be no discontinuities at the VLBW cutoff.[18] Table 1 provides the results. Panels A and B use the FC sample and check whether maternal and focal child characteristics are balanced, while Panels C and D use the sibling sample to check for discontinuities in the covariates of siblings.[19] Columns 1-5 report results for (siblings of) focal children with gestational age of at least 32 weeks and Columns 6-10 for those with gestational age of less than 32 weeks. The results show that observations just below the VLBW cutoff are generally similar to those just above the VLBW cutoff in terms of maternal characteristics, focal child characteristics, and sibling characteristics. There are few characteristics that are imbalanced across the threshold and only marginally so (e.g., immigration status of the mother and birth weight of the sibling). In order to check whether such imbalance matters in the specific context of our outcomes, we investigate whether predicted outcomes based on observable characteristics are smooth across the cutoff. In particular, we predict each outcome variable using

[17] The estimates corresponding to Figures 1(a)-(b) are 0.196 (bias-corrected estimate -12.614, s.e. 17.429) and -1.765 (bias-corrected estimate -2.645, s.e. 9.176). The results are robust to using the logarithm of the number of births as the dependent variable instead. In this case, the estimated coefficients are 0.027 (bias-corrected estimate -0.238, s.e. 0.324) and -0.008 (bias-corrected estimate -0.002, s.e. 0.175).

[18] Visual evidence from selected covariates is provided in the Appendix. Appendix Figures A2-A3 present means in the sibling sample by birth weight of the focal child, separately for siblings of focal children with gestational age above and below 32 weeks. Appendix Figures A4-A5 plot the distribution of selected observable characteristics in the FC sample for focal children with gestational age above and below 32 weeks, respectively.

[19] The covariate tests are based on the full sibling sample. Tests based on the subsamples of siblings for whom we have test score or enrollment information yield similar results (available upon request).

a linear model including the full set of control variables.[20] As the last panel of the Table shows, there is no significant discontinuity in any of the predicted outcomes across the cutoff.

Overall, the analyses in this section indicate that there is no evidence of manipulation of the running variable around the VLBW cutoff or of discontinuities in the observable characteristics of focal children, their mothers and their siblings.

6.2. Baseline Results

Figures 2-3 provide visual evidence on the relationship between focal child birth weight and the academic outcomes of their siblings.[21] Since focal children with a gestational age of less than 32 weeks are eligible to receive medical treatments regardless of their birth weight, we plot the distribution of outcomes separately by the gestational age of focal children. Any discontinuity in the outcomes of siblings of focal children with less than 32 weeks of gestational age would suggest a violation of the key identification assumptions underlying the RD design.

Figures 2(a) and 2(c) show that siblings of GA32+ focal children with birth weight slightly lower than 1,500 grams have visibly higher test scores in both language and math. Distributions of test scores, on the other hand, are relatively smooth across the VLBW threshold for siblings of GA32- focal children as shown in Figures 2(b) and 2(d). Figure 3, on the other hand, does not indicate important spillovers for enrollment outcomes.

In Table 2, we present the corresponding regression results from our baseline models. We again present our findings separately by gestational age of focal children. Each cell reports the estimated coefficient of $VLBW$ from a different regression. Consistent with the graphical evidence, we find strong evidence of positive spillovers on test scores.[22] For example, siblings of VLBW newborns with gestational age of at least 32 weeks have 9th grade language (math) test scores that are on average 0.386 (0.255) standard deviations higher.[23] In contrast, the results indicate that the siblings

[20] We use the universe of births when predicting the outcomes.

[21] All figures plotting raw data use 25-gram bins to reduce noise.

[22] Among test-takers in the sibling sample, the maximum age difference between older siblings and focal children is 7.5 years, meaning that none of the older siblings take the test before the focal children are born.

[23] The 95% robust confidence intervals are constructed as: $0.478 \pm 1.96 \cdot 0.199 = [0.087, 0.867]$ for language test scores and $0.255 \pm 1.96 \cdot 0.180 = [0.062, 0.768]$ for math test scores.

of focal children with gestational age below 32 weeks have similar test scores across the VLBW threshold.

The estimated effects are economically significant. One way to gauge their magnitude is by looking at other policy-relevant test score gaps. For example, among all children born during the period covered by our sibling sample, the difference in language (math) scores between the children of non-immigrants and immigrants is 0.264 (0.404) standard deviations. Our results imply that medical interventions are equivalent to eliminating the language disadvantage for children of immigrants and reducing the gap in math scores by more than half. We also calculate that the difference in language (math) test scores among those born in households above the 90th income percentile and those born in households below the 10th income percentile is 0.557 (0.769) standard deviations. Our coefficients suggest that medical interventions can reduce the income-based test score gap at age 16 by 33-69%. These effects are in line with those found by Duncan and Sojourner (2003) for income-based test score gaps at ages 3 through 8 for children exposed to an early-education program targeting low-birth-weight children in the US.

Despite the strong spillovers on test scores, it does not appear that there are significant spillover effects on the likelihood of continuing education beyond compulsory schooling. This is likely due to the high rate of high school enrollment in the sample (78%). While we find some weak evidence of positive effects on enrollment in an academic track and negative effects on enrollment in a vocational track, these estimates are sensitive to sample and model specification. For this reason, in the rest of the paper we focus on test scores for which we find much stronger evidence of spillover effects.

6.3. Robustness Checks

In this section we present robustness checks using the GA32+ sibling sample.[24] Appendix Table A3 and Figure 4 investigate the robustness of our estimates to the choice of bandwidth and degree of polynomial in the running variable. We present results for all bandwidths between 100-300 grams in 10-gram steps. For each bandwidth, we provide results using up to a second-degree polynomial in birth weight. The Figure shows that the magnitudes of the estimates are remarkably consistent across different bandwidths, regardless of the degree of polynomial in the running variable.

[24] Appendix Tables A5-A7 and Figure A6 present corresponding results from the GA32- sibling sample.

Table 3 provides additional sensitivity analyses. In Column 1, we check the sensitivity of the results to the inclusion of the control variables described in Section 5. If the key assumption in our RD design is satisfied (i.e., birth weight is as good as random around the cutoff), then including additional covariates should not impact the estimates but only increase precision. The results show that this is indeed the case: siblings of focal children who were slightly below the cutoff have significantly higher 9th grade test scores and the magnitudes of the effects are very similar to those in the baseline with slightly smaller standard errors.

Columns 2-4 turn to the role of heaping. Following Barreca et al. (2011), our main specification controls for heaping at 50-gram intervals. In Column 2, we check whether our results are robust to excluding heaping dummies. Given that our data records birth weight in grams and heaping is generally symmetric around the cutoff, we expect heaping to be less of a concern in our context. The estimated coefficients in Column 2 confirm this prior. We also implement an alternative method suggested by Barreca et al. (2011) and estimate “donut” regressions that exclude observations close to the cutoff. In Column 3, we exclude siblings of focal children who weighed 1,500 grams, while in Column 4 we further exclude siblings of focal children with birth weight between 1,490 to 1,510 grams. The results are again similar to the main estimates, suggesting that our baseline results are not driven by heaping.

In Columns 5-8, we investigate the sensitivity of our results to model specification. Our baseline model uses a triangular kernel. We show that our findings are robust to using a rectangular kernel that places equal weights to each observation (Column 5). In Column 6, we allow the bandwidths to differ across outcome variables, using the optimal bandwidths suggested by the Calonico et al. (2014) strategy. Given the stability of the estimates to alternative bandwidths, it is not surprising that the results are again very robust. Columns 7-8 check the sensitivity of our inference by clustering standard errors at the birth weight (Column 7) or mother (Column 8) level. In both cases, the results remain statistically significant at conventional levels.

In Table 4 we check the robustness of our results to sample selection. To the extent that the birth weight of children is correlated within the family, it may be that siblings of VLBW children are more likely to be VLBW themselves. If this is the case, then the observed academic achievement gains among siblings may be due to the early-life medical interventions they themselves received at birth instead of spillovers from the treatments of their siblings. In order to shed light on this issue, we exclude VLBW siblings (Column 1) and confirm that our main results are not driven by them.[25]

Multiple births are generally characterized by lower birth weight. Indeed, multiple births represent a disproportionate share of focal children within our bandwidth relative to their share in the full population of births (18.11 percent vs. 2.37 percent). But multiple births may also impact siblings through channels other than medical treatments, such as family size. Therefore, Column 2 investigates the robustness of our results in a sample of siblings of singleton focal children. We confirm that our baseline results are not sensitive to this sample restriction. This should not be surprising since we do not find any discontinuity in the probability of a multiple birth across the VLBW threshold (see Table 1).

Previous literature finds that early-life medical treatments have significant effects on focal child survival and we confirm these results in our context in the Appendix. This means that the spillover effects to siblings may also be due to changes in family size. In Column 3 we check if our baseline results still hold when we restrict the sample to siblings of focal children who survive past the first year of life. The results are similar to the baseline with slightly larger magnitudes, indicating again that our results are not due to differences in family size across the VLBW cutoff. In Column 4 we investigate the role of multiple births and focal child survival jointly, focusing only on the sample of singleton siblings and singleton focal children who survive to the birth of the sibling. Our results again indicate significant spillovers to siblings.

Bharadwaj et al. (2013) note that being very low birth weight may signal different underlying health issues for GA32+ and GA32- focal children, potentially resulting in different medical treatments at the cutoff. This would create a challenge when using siblings of GA32- focal children as a falsification test. In order to increase the comparability of the two samples, we restrict the sample to siblings of GA32+ focal children with 32 and 33 weeks of gestation and to siblings of GA32- focal children with gestational age of 30 and 31 weeks. The results in Column 5 of Table 4 and of Appendix Table A7 confirm strong positive spillovers to siblings in the GA32+ sample and no spillovers to siblings in the GA32- sample.

[25] After excluding VLBW siblings, only 10 siblings with a gestational age below 32 weeks remain in the sample. Further dropping these from the sample does not change the results.

Finally, we conduct two falsification checks. First, we investigate the presence of discontinuities in 28-day and 1-year mortality of older siblings. If families of VLBW children are prone to having health shocks and our human capital achievement results capture the effects of unobserved family traits instead of spillovers from early-life medical interventions, then we may expect to see differential survival rates for older siblings who were not exposed to VLBW focal children. Using sibling mortality rates as outcomes indicates that this is not a concern in our context (see Panel D in Appendix Table A2).

Second, we check whether we observe similar improvements in the test scores of siblings at other points in the distribution of birth weight of the focal child. If the observed gains in academic achievement are indeed driven by the medical treatments received by focal children, then we should not observe systematic discontinuities in the educational outcomes of siblings at other potential cutoffs. We examine cutoffs from 1,100 grams to 3,100 grams, keeping the bandwidth fixed at 200 grams on either side of the cutoff. The results presented in Appendix Table A4 and Figure 5 indicate that there is no other cutoff where either of the test scores exhibit gains of a magnitude comparable to those observed at the 1,500 gram cutoff.[26] Combined with the absence of discontinuities at the VLBW cutoff in the educational outcomes of siblings of focal children with gestational age of less than 32 weeks, these findings strongly suggest that the observed spillover effects are due to the impact of medical treatments provided to VLBW focal children.

6.4. Potential Mechanisms

We now attempt to investigate the potential channels described in Section 3. We first examine whether improved sibling academic outcomes could result from improved sibling health. In particular, we study whether the early-life health interventions provided to VLBW focal children affect the health outcomes of their siblings as proxied by hospital admissions and ER visits in five-year intervals after the birth of the focal child. The results in Table 5 do not show any improvement in the physical health of siblings up to 10 years after the birth of the focal child. Hence, the observed spillover effects are unlikely to be driven by siblings' direct or indirect exposure to additional medical care.

[26] There are only two other statistically significant coefficients: for language test scores at 2,500 grams and for math test scores at 2,900 grams. The magnitudes of these effects are three to four times smaller than the estimated effects at 1,500 grams.

As discussed before, previous studies document that early-life health interventions improve both short-term health and long-term academic outcomes of focal children. In Appendix Table A8, we replicate these findings using the FC sample. Consistent with the previous literature, we find in the GA32+ sample that the probability of death within the first 28 days (1 year) of life is 4.1 (5.4) percentage points lower among VLBW newborns. These are large gains when compared to the average mortality rates of those above the cutoff (6.2 and 7.7 percent, respectively) but they are comparable in magnitude to the reductions in infant mortality from previous studies: 1 percentage point (mean: 5.5 percent) in the US (Almond et al., 2010); 4.5 percentage points (mean: 11 percent) in Chile and 3.1 percentage points (mean: 3.6 percent) in Norway (Bharadwaj et al., 2013). We also show that focal children who were just below the VLBW cutoff have better academic achievement in the long run, with 9th grade language and math test scores higher on average by 0.229 and 0.315 standard deviations, respectively.[27] Given the previous studies documenting spillover effects of a disabled sibling, we also investigate whether the medical treatments provided to VLBW children affect their likelihood of having a disability. Our results do not indicate any discontinuity at the cutoff in the likelihood of being diagnosed by age 10 with a range of common childhood disabilities: ADHD, intellectual disability, behavioral and emotional disorders, epilepsy, or cerebral palsy (see Panel C of Appendix Table A8). However, we do find some evidence in support of improved long-term health during school age years, as proxied by hospital admissions and ER visits (see Panel D of Appendix Table A8).

In Table 6, we examine whether these improved focal child outcomes are accompanied by changes in family resources. We construct measures of parental and total income as well as parental labor market participation (an indicator for being employed at least one day during the year and average number of full-time working days per year) in five-year intervals after the birth of the focal child. We generally do not find significant discontinuities at the VLBW cutoff in measures of family resources, except for some evidence suggesting improved labor market outcomes for fathers 6-10 years after the birth of the focal child. However, this does not translate into higher total family income during the same period. Therefore, we interpret the evidence in Table 6 to suggest that

[27] The estimated effect on math test scores is comparable to those found by Bharadwaj et al. (2013), who estimate effects of 0.152 standard deviations in Chile and 0.476 standard deviations in Norway. These results are not driven by delayed school entry as proxied by the age at which focal children take the 9th grade test (Landersø et al., 2017), as shown in Panel B of Appendix Table A2.

differences in total household resources (both time and money) are unlikely to explain the observed spillover effects on siblings.

We also study whether early-life medical treatments to VLBW children are associated with changes in the family environment. Motivated by the literature linking child health to family dissolution and parental health, we check in Table 7 if there are any discontinuities across the VLBW threshold in terms of divorce and parental mental health as proxied by the use of antidepressants. We find no significant difference in the likelihood of family dissolution across the VLBW cutoff up to ten years after the birth of the focal child. However, we do find some evidence of improved maternal mental health soon after the birth of the focal child that dissipates as the child ages.[28] To the extent that better mental health leads to better parent-child interactions, this could be one of the main channels behind our results.

In order to shed some light on the quality of sibling interactions, we examine the existence of spillover effects in subsamples defined by sibship characteristics. Previous literature in psychology and in economics finds that girls, younger siblings, and siblings of the same sex are more likely to be affected by the interaction with their siblings (e.g., Furman and Buhrmester, 1985; Dunn, 2007; Oettinger, 2000; Fletcher et al., 2012). In addition, the peer effects literature in economics suggests that peer effects on math test scores last longer than on language test scores (e.g., Neidell and Waldfogel, 2010). Although we are somewhat underpowered, we find suggestive evidence in line with these studies in Table 8. In particular, we document larger spillover effects on the math test scores of girls and of siblings of the same sex, although we cannot reject the equality of effects across subsamples.[29] We cautiously interpret these results as evidence that improved quality of sibling interactions may be one of the drivers of improved sibling academic achievement.

---

[28] The improvements in maternal mental health are also present in the sample of focal children who survive past the first year of life, suggesting that they are not due to maternal grief.

[29] The sample of older siblings is too small for meaningful comparisons with younger siblings. We also split the sibling sample by the median age difference with respect to the focal child and find relatively similar spillovers to closely-spaced siblings and to siblings born more than 3.5 years apart (see Appendix Table A9). Finally, we estimate specifications similar to our baseline but where we control for the corresponding focal child test score. This strategy wipes out the discontinuity in sibling math test scores at the cutoff. Although the estimates are not causal, they do suggest again that improvements in the "quality" of focal children may be an important driver of observed spillovers on sibling test scores (results available upon request).

Finally, changes in the intra-household allocation of resources, particularly parental compensating behavior, may be another mechanism behind our results. Data limitations do not allow us to investigate this channel directly, but we can provide some indirect evidence if we make the assumption that there are dynamic complementarities in the production of human capital, as suggested by Cunha and Heckman (2007). In this case, children with high initial endowments would benefit most from parental investments because "skills beget skills." Consider two children with low initial endowment, A and B, who are identical in every respect except that A has a sibling with birth weight slightly below the VLBW cutoff while B has a sibling with birth weight slightly above the cutoff. If both sets of parents engage in compensating behavior, then child B has more resources taken away from her and allocated to her sibling than child A does (because the VLBW sibling of child A benefits from the additional medical treatments). Therefore, in the long term child B ends up with a lower level of skills than child A. Now consider a similar pair of identical children, C (who has a VLBW sibling) and D (who does not), but with high initial endowment. Just as before, child D has more resources taken away from her and so she ends up with a lower level of skills in the long term than child C. However, because of dynamic complementarities, child D is harmed even more by the fewer resources she receives because the return to those resources would be higher for her than for child B. Therefore, the difference in skills between children C and D (high initial endowment) is larger than the difference in skills between children A and B (low ability).[30]

To check whether we observe this pattern in our data, we rely on birth weight as an indicator of initial endowments because the previous literature finds that it is is highly correlated with later-life academic, health, and labor market outcomes (e.g., Black et al., 2007; Figlio et al., 2014). We define "high endowment" siblings as those whose birth weight is higher than the birth weight of the median child born during our sample period. The results shown in Appendix Table A10, although relatively imprecise, suggest that high-endowment siblings benefit more than low-endowment siblings from the additional medical treatments received by VLBW focal children. This suggests that parental compensating behavior (possibly combined with dynamic

[30] Alternatively, the difference in skills between children C and D can be larger than the difference in skills between children A and B if parents reallocate more resources to the focal child in order to compensate for the larger difference in endowments within the family. This explanation is also consistent with compensating behavior by the parents.

complementarities in the production of human capital) may be one of the factors behind the observed spillover effects.

7. Conclusions

The emerging economic literature on the importance of sibling interactions has prompted the question of whether interventions that improve child health can affect these spillover effects. We answer this question by investigating the spillover effects of early-life medical treatments provided to VLBW children on the human capital accumulation of their siblings. Using register data from Denmark, we find that siblings of focal children who were slightly below the VLBW cutoff have better 9th grade language and math test scores. Our results suggest that improved quality of parent-child and sibling interactions and parental compensating behavior may be important in explaining these effects.

Our results underscore the importance of health interventions targeted to other family members, specifically siblings, as an important factor in the accumulation of human capital. Our findings also have important implications for understanding the efficacy of early-life medical interventions. In particular, they underline the need to consider potential externalities when assessing the net benefits of medical treatments. Finally, our results have implications for studies on the effects of early-life health endowments using sibling fixed-effects estimators. The fact that we find substantial positive spillovers on the siblings of treated children suggests that within-sibling comparisons of achievement gains may underestimate the true impact of initial health endowments on later-life outcomes.

References

Adhvaryu, Achyuta, and Anant Nyshadham. 2016. "Endowments at Birth and Parents' Investments in Children." *Economic Journal* 126 (593): 781-820.

Almond, Douglas, and Janet Currie. 2011. "Human Capital Development before Age Five." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, Vol. 4B. Amsterdam; New York; New York, N.Y., U.S.A.: North Holland.

Almond, Douglas, Janet Currie, and Valentina Duque. Forthcoming. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*.

Almond, Douglas, Joseph J. Doyle, Amanda E. Kowalski, and Heidi Williams. 2010. "Estimating Marginal Returns to Medical Care: Evidence from At-Risk Newborns." *Quarterly Journal of Economics* 125 (2): 591–634.

Almond, Douglas, and Bhashkar Mazumder. 2013. "Fetal Origins and Parental Responses." *Annual Review of Economics* 5 (1): 37–56.

Altonji, Joseph G., Sarah Cattan, and Iain Ware. 2016. "Identifying Sibling Influence on Teenage Substance Use." *Journal of Human Resources* 52(1): 1-47.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. 2011. "Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification." *Quarterly Journal of Economics* 126 (4): 2117–23.

Behrman, Jere R., Robert A. Pollak, and Paul Taubman. 1982. "Parental Preferences and Provision for Progeny." *Journal of Political Economy* 90 (1): 52–73.

Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman. 1994. "Endowments and the Allocation of Schooling in the Family and in the Marriage Market: The Twins Experiment." *Journal of Political Economy* 102 (6): 1131–74.

Bharadwaj, Prashant, Juan Eberhard, and Christopher Neilson. Forthcoming. "Health at Birth, Parental Investments and Academic Outcomes." *Journal of Labor Economics*.

Bharadwaj, Prashant, Katrine Vellesen Løken, and Christopher Neilson. 2013. "Early Life Health Interventions and Academic Achievement." *American Economic Review* 103 (5): 1862–91.

Bütikofer, Aline, Katrine V. Løken, and Kjell Salvanes. Forthcoming. "Infant Health Care and Long Tern Outcomes," *Review of Economics and Statistics*.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Roció Titiunik. 2018. "Regression Discontinuity Designs Using Covariates." Mimeo.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Chay, Kenneth Y., Jonathan Guryan, and Bhashkar Mazumder. 2009. "Birth Cohort and the Black-White Achievement Gap: The Roles of Access and Health Soon After Birth." NBER Working Paper No. 15078.

Corman, Hope, and Robert Kaestner. 1992. "The Effects of Child Health on Marital Status and Family Structure." *Demography* 29 (3): 389–408.

Corman, Hope, Kelly Noonan, and Nancy E. Reichman. 2005. "Mothers' Labor Supply in Fragile Families: The Role of Child Health." *Eastern Economic Journal* 31 (4): 601–16.

Cunha, F., and J. Heckman. 2007. "The technology of skill formation." *American Economic Review* 97 (2): 31–47.

Cunha, F., J. J. Heckman, and S. M. Schennach. 2010. "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica* 78 (3): 883–931.

Cutler, David M., and Ellen Meara. 1998. "The Medical Costs of the Young and Old: A Forty-Year Perspective." In *Frontiers in the Economics of Aging*, edited by David A. Wise, 215–46. University of Chicago Press.

Dahl, Gordon, Katrine V. Løken, and Magne Mogstad. 2014. "Peer Effects in Program Participation." *American Economic Review* 104: 2049–2074.

Daysal, N. Meltem, Mircea Trandafir, and Reyn van Ewijk. 2015. "Saving lives at birth: The impact of home births on infant outcomes." *American Economic Journal: Applied Economics* 7(3): 1-24.

Daysal, N. Meltem, Mircea Trandafir, and Reyn van Ewijk. Forthcoming. "Low-risk isn't no-risk: Perinatal treatments and the health of low-income newborns." *Journal of Health Economics*.

Duncan, Greg J. and Aaron J. Sojourner. 2013. "Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps?" Journal of Human Resources 48(4), 945–68.

Dunn J. 2007. "Siblings and Socialization." In *Handbook of Socialization: Theory and Research* edited by JE Grusec and PD Hastings PD, 309–327. New York: Guilford Press.

Field, Erica, Omar Robles, and Maximo Torero. 2009. "Iodine Deficiency and Schooling Attainment in Tanzania." *American Economic Journal: Applied Economics* 1 (4): 140–69.

Fletcher, J., Nicole L. Hair, and Barbara L. Wolfe. 2012. "Am I my brother's keeper? Sibling spillover effects: The case of developmental disabilities and externalizing behavior." NBER WP #18279.

Furman, Wyndol, and Duane Buhrmester. 1985. "Children's Perceptions of the Qualities of Sibling Relationships." *Child Development* 56 (2): 448.

Ginther, Donna K., and Robert A. Pollak. 2004. "Family Structure and Children's Educational Outcomes: Blended Families, Stylized Facts, and Descriptive Regressions." *Demography* 41 (4): 671–96.

Greisen, G, M. B. Petersen, S. A. Pedersen, and P. Bekgaard. 1986. “Status at Two Years in 121 Very Low Birth Weight Survivors Related to Neonatal lntraventricular Haemorrhage and Mode of delivery”. *Acta Paediatrica Scandinavica* 75: 24-30.

Haveman, Robert, and Barbara Wolfe. 1995. “The Determinants of Children’s Attainments: A Review of Methods and Findings.” *Journal of Economic Literature* 33 (4): 1829–78.

Health Care in Denmark. 2008. Danish Ministry of Health and Prevention (“Ministeriet for Sundhed og Forbyggelse”).

Hertz, Birgitte, Eva-Bettina Holm, and Jørgen Haahr. 1994. “Prognosen for børn med meget lav fødselsvægt i Viborg Amt.” *Ugeskrift for læger* 156 (46): 6865−6868.

Hjort, Jonas, Mikkel Sølvsten, and Miriam Wüst. 2017. “Universal Investment in Infants and Long-run Health: Evidence From Denmark’s 1937 Home Visiting Program.” *American Economic Journal: Applied Economics* 9(4): 78-104.

Jacobsen, Thorkild, John Grønvall, Sten Petersen, and Gunnar Eg Andersen. 1993. “‘Minitouch’ treatment of very low-birth-weight infants.” *Acta Paediatrica Scandinavica* 82: 934-8.

Joensen, Juanna and Helena Skyt Nielsen. 2018. “Spillovers in Educational Choice.” *Journal of Public Economics* 157: 158-83.

Johnson, T. J., Patel, A. L., Jegier, B. J., Engstrom, J. L., & Meier, P. P. 2012. Cost of morbidities in very low birth weight infants. *The Journal of Pediatrics*, 162(2), 243-49.e1.

Jürges, Hendrik, and Juliane Köberlein. 2015. “What explains DRG upcoding in neonatology? The roles of financial incentives and infant health.” *Journal of Health Economics* 43: 13-26.

Kvist, Anette Primdal, Helena Skyt Nielsen, and Marianne Simonsen. 2013. “The Importance of Children’s ADHD for Parents’ Relationship Stability and Labor Supply.” *Social Science & Medicine* 88 (July): 30–38.

Landersø, Rasmus, Helena Skyt Nielsen, and Marianne Simonsen. 2017. “School Starting Age and the Crime-Age Profile.” *Economic Journal* 127: 1096-111.

Levav, I, R Kohn, J Iscovich, J H Abramson, W Y Tsai, and D Vigdorovich. 2000. “Cancer Incidence and Survival Following Bereavement.” *American Journal of Public Health* 90 (10): 1601–7.

Li, Jiong, Thomas Munk Laursen, Dorthe Hansen Precht, Jørn Olsen, and Preben Bo Mortensen. 2005. “Hospitalization for Mental Illness among Parents after the Death of a Child.” *New England Journal of Medicine* 352 (12): 1190–96.

Li, Jiong, Dorthe Hansen Precht, Preben Bo Mortensen, and Jørn Olsen. 2003. "Mortality in Parents after Death of a Child in Denmark: A Nationwide Follow-up Study." *Lancet* 361 (9355): 363–67.

Manski, Charles F., Gary D. Sandefur, Sara McLanahan, and Daniel Powers. 1992. "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation." *Journal of the American Statistical Association* 87 (417): 25–37.

Mathiasen, René, Bo M. Hansen, Anne Løkke, and Gorm Greisen. 2008. "Treatment of Preterm Children at Rigshospitalet during the period 1955-2007" (Behandling af tidligt fødte børn på Rigshospitalet i perioden 1955–2007). *Bibliotek for Læger* 200 (4): 528–46.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Neidell, Matthew, and Jane Waldfogel. 2010. "Cognitive and Noncognitive Peer Effects in Early Education." *Review of Economics and Statistics* 92 (3): 562–76.

Newhouse, Joseph P. 1992. "Medical Care Costs: How Much Welfare Loss?" *Journal of Economic Perspectives* 6 (3): 3–21.

Nicoletti, Cheti and Birgitta Rabe. Forthcoming. "Sibling spillover effects in school achievement." *Journal of Applied Econometrics*.

Noonan, Kelly, Nancy E. Reichman, and Hope Corman. 2005. "New Fathers' Labor Supply: Does Child Health Matter?" *Social Science Quarterly* 86 (December): 1399–1417.

Oettinger, Gerald S. 2000. "Sibling Similarity in High School Graduation Outcomes: Causal Interdependency or Unobserved Heterogeneity?" *Southern Economic Journal* 66 (3): 631–48.

Ouyang, Lijing. 2005. "Three Essays on Teen Risky Behaviors." Ph.D. dissertation, Duke University.

Parman, John. 2015. "Childhood Health and Sibling Outcomes: Nurture Reinforcing Nature during the 1918 Influenza Pandemic." *Explorations in Economic History* 58 (October): 22–43.

Peitersen, Birgit and Mette Arrøe. 1991. "Neonatologi – Det raske og det syge nyfødte barn". Nyt Nordisk Forlag Arnold Busck.

Pitt, Mark M., Mark R. Rosenzweig, and Md. Nazmul Hassan. 1990. "Productivity, Health, and Inequality in the Intrahousehold Distribution of Food in Low-Income Countries." *American Economic Review* 80 (5): 1139–56.

Powers, Elizabeth T. 2003. "Children's Health and Maternal Work Activity." *Journal of Human Resources* 38 (3): 522–56.

Reichman, Nancy E., Hope Corman, and Kelly Noonan. 2004. “Effects of Child Health on Parents’ Relationship Status.” *Demography* 41 (3): 569–84.

Rosenzweig, Mark R., and Kenneth I. Wolpin. 1988. “Heterogeneity, Intrafamily Distribution, and Child Health.” *Journal of Human Resources* 23 (4): 437–61.

Shigeoka, Hitoshi, and Kiyohide Fushimi. 2014. “Supplier-Induced Demand for Newborn Treatment: Evidence from Japan.” *Journal of Health Economics* 35 (May): 162–78.

Schiøtz, Peter Oluf and Skovby, Flemming. (2001). Praktisk pædiatri, Munksgaard Danmark.

Thomsen, Ketty Dahl, Helle Hansen, Finn Ebbesen, and Vibeke Jacobsen. 1991. “Neonatal mortalitet hos børn med meget lav fødselsvægt i Nordjyllands Amt: en retrospektiv opgørelse.” *Ugeskrift for læger* 153 (47): 3310−3313.

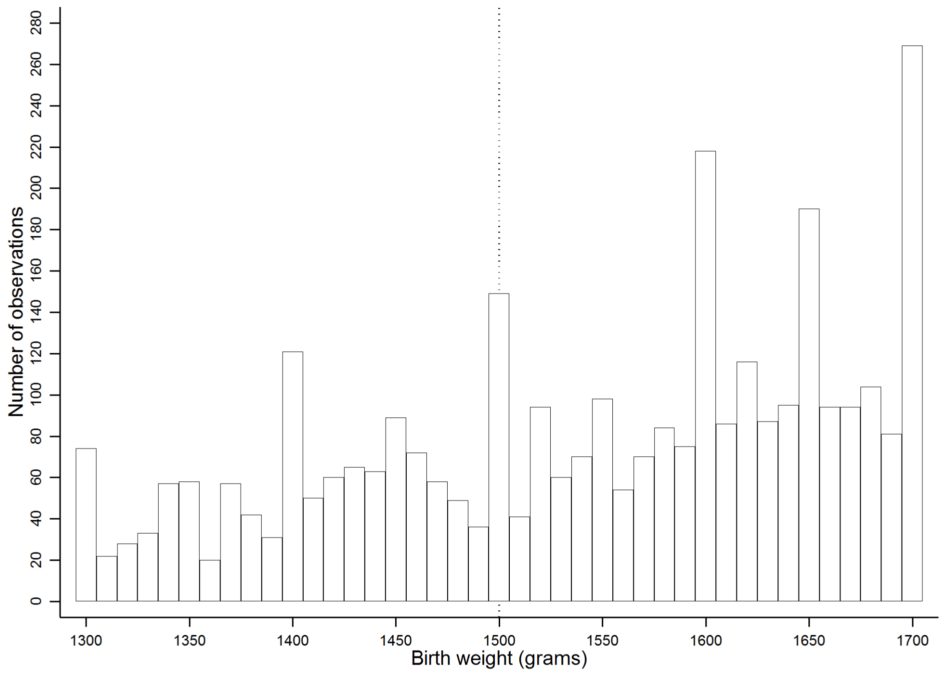
Topp, Monica, Peter Uldall, and Gorm Greisen. 2001. “Cerebral palsy births in Eastern Denmark, 1987–90: implications for neonatal care.” *Paediatric and Perinatal Epidemiology* 15: 271–277.

Verder, Henrik. 2007. “Nasal CPAP has become an indispensable part of the primary treatment of newborns with respiratory distress syndrome.” *Acta Pædiatrica* 96: 482–484.

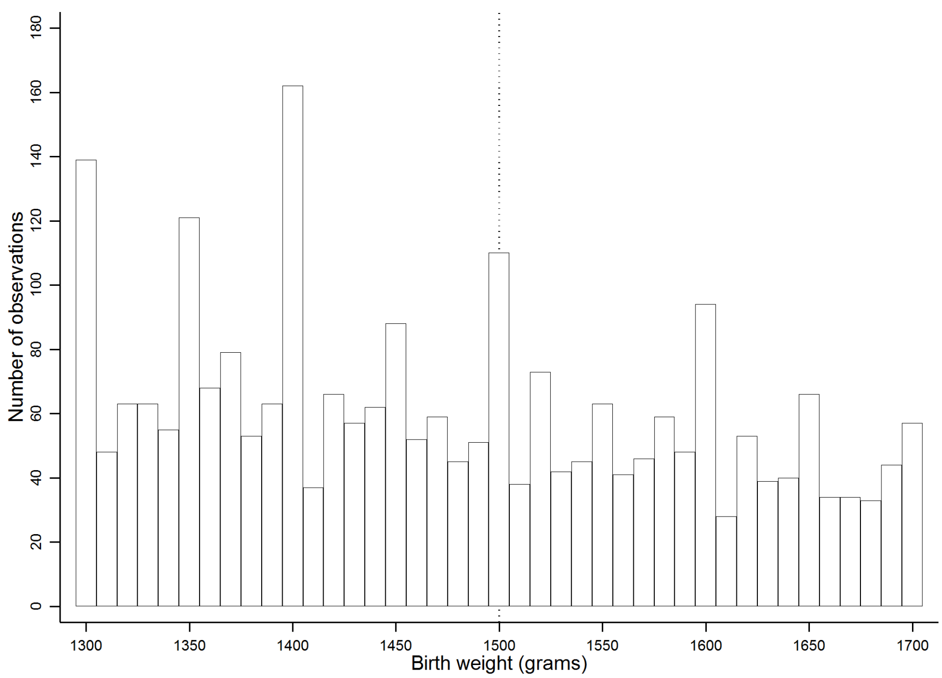
Verder, Henrik, Bengt Robertson, Gorm Greisen, Finn Ebbesen, Per Albertsen, Kaare Lundstrøm, and Thorkild Jacobsen. 1994. “Surfactant Therapy and Nasal Continuous Positive Airway Pressure for Newborns with Respiratory Distress Syndrome.” *New England Journal of Medicine* 331 (16): 1051-1055.

Wasi, Nada, Bernard van den Berg, and Thomas C. Buchmueller. 2012. “Heterogeneous Effects of Child Disability on Maternal Labor Supply: Evidence from the 2000 US Census.” *Labour Economics* 19 (1): 139–54.

Yi, Junjian, James Heckman, Junsen Zhang, and Gabriella Conti, 2015. “Early Health Shocks, Intra-Household Resource Allocation and Child Outcomes.” *Economic Journal* 125: F347-F371.

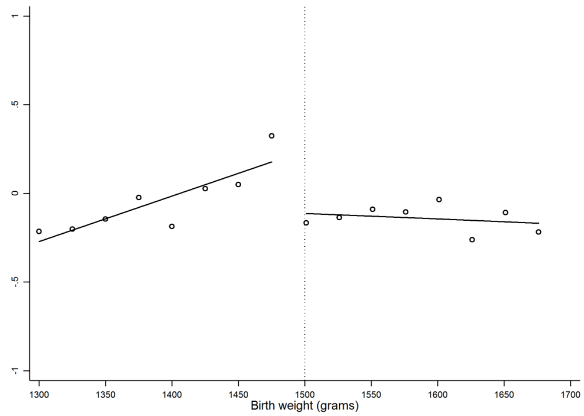


(a) Siblings of focal children with gestational age ≥ 32 weeks

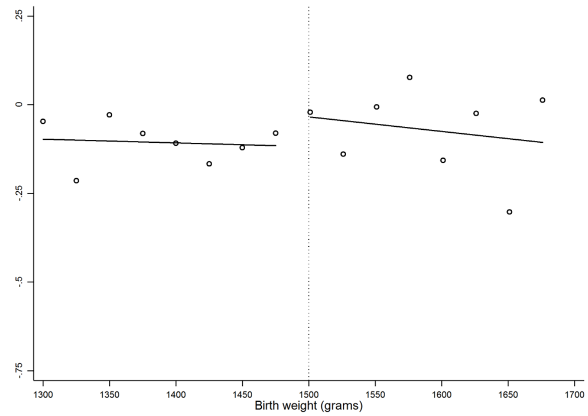


(b) Siblings of focal children with gestational age < 32 weeks

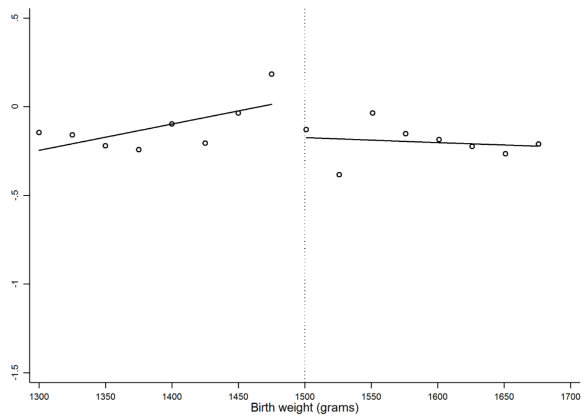
Figure 1: Frequency of observations around the VLBW cutoff, sibling sample



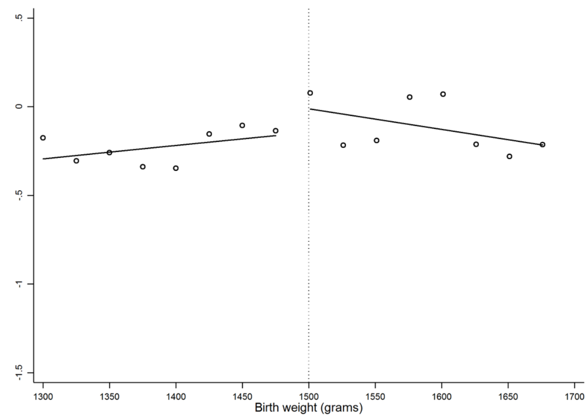
(a) Language test score, GA32+



(b) Language test score, GA32-



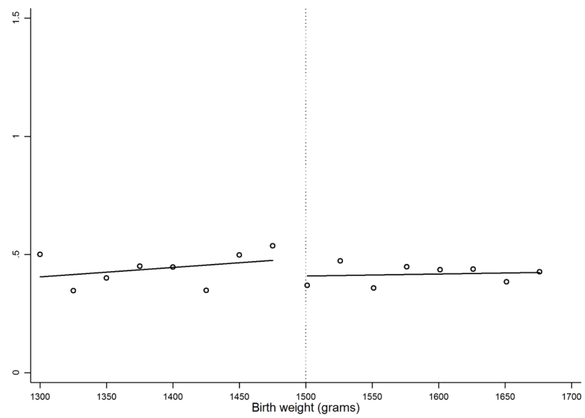
(c) Math test score, GA32+



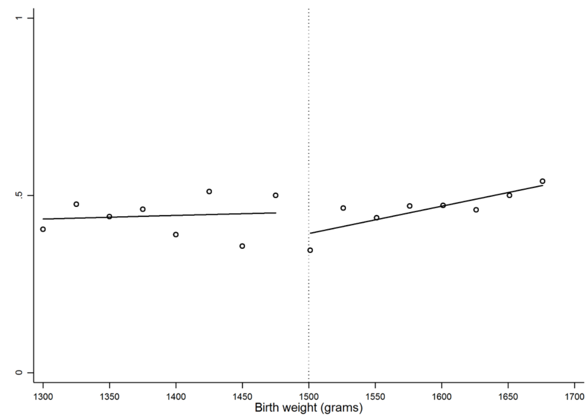
(d) Math test score, GA32-

Notes: Each dot represents the average of the variable indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first-degree polynomial estimated separately on either side of the VLBW cutoff.

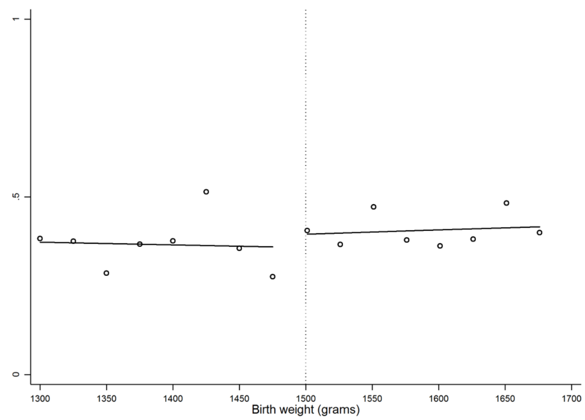
Figure 2: Distribution of sibling test scores around VLBW cutoff



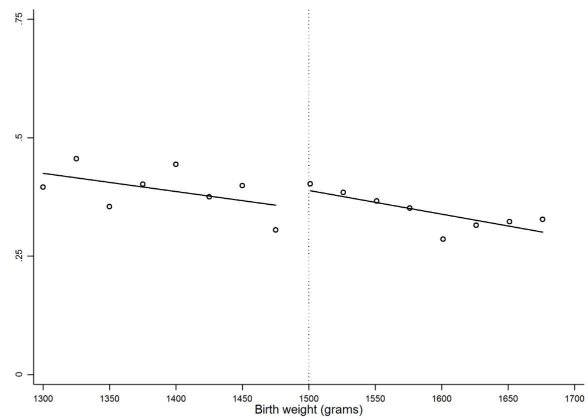
(a) Enrollment in academic track, GA32+



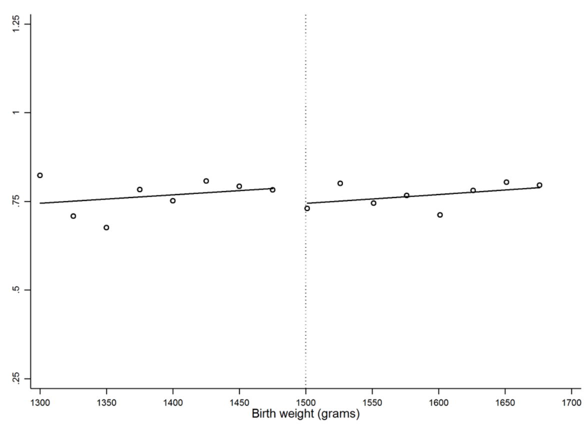
(b) Enrollment in academic track, GA32-



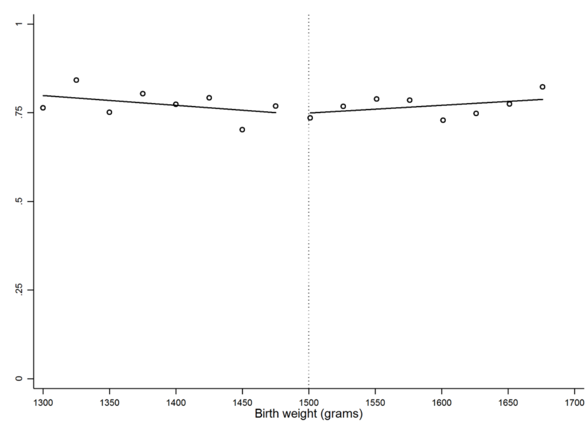
(c) Enrollment in vocational track, GA32+



(d) Enrollment in vocational track, GA32-



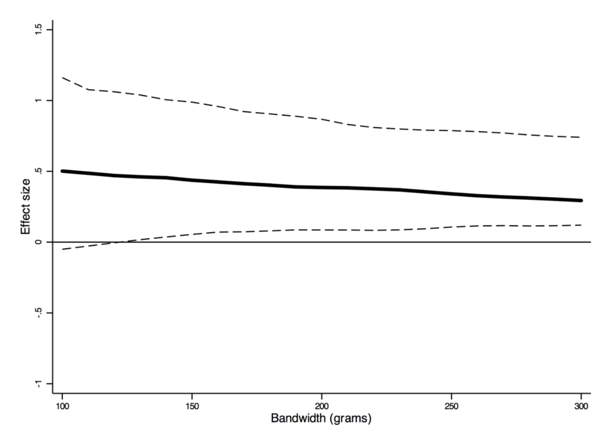
(e) Enrollment beyond compulsory schooling, GA32+



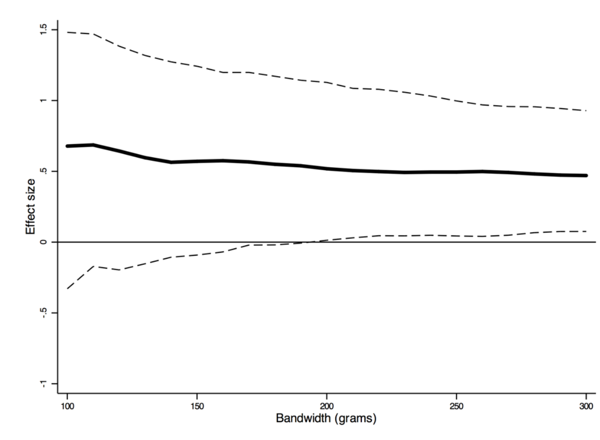
(f) Enrollment beyond compulsory schooling, GA32-

Notes: Each dot represents the average of the variable indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first-degree polynomial estimated separately on either side of the VLBW cutoff.

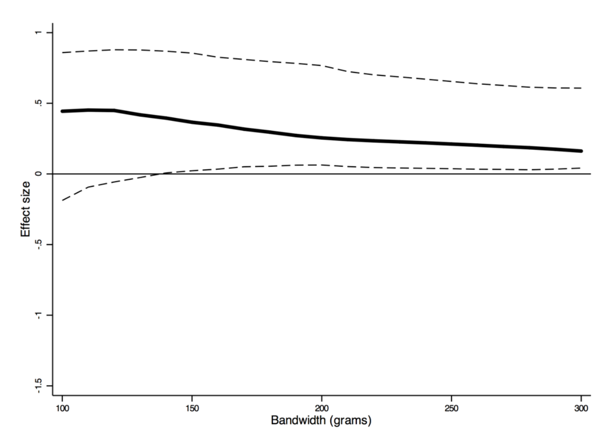
Figure 3: Distribution of sibling enrollment outcomes around VLBW cutoff



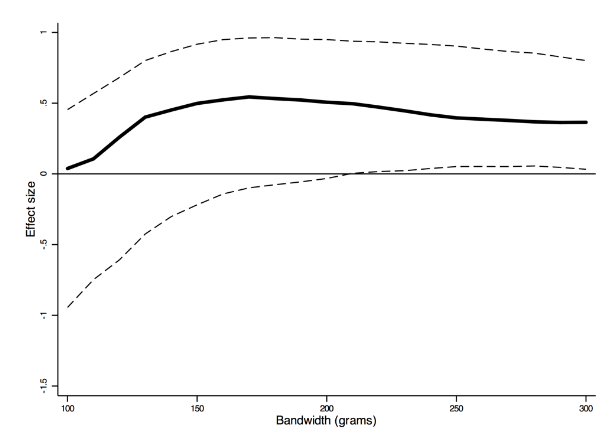
(a) Language test score, first degree polynomial



(b) Language test score, second degree polynomial



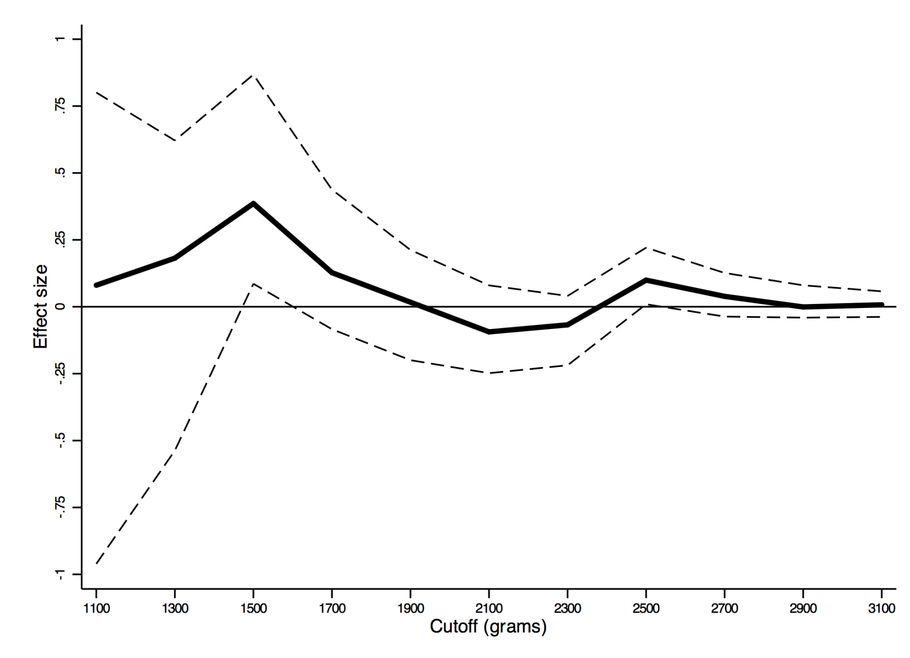
(c) Math test score, first degree polynomial



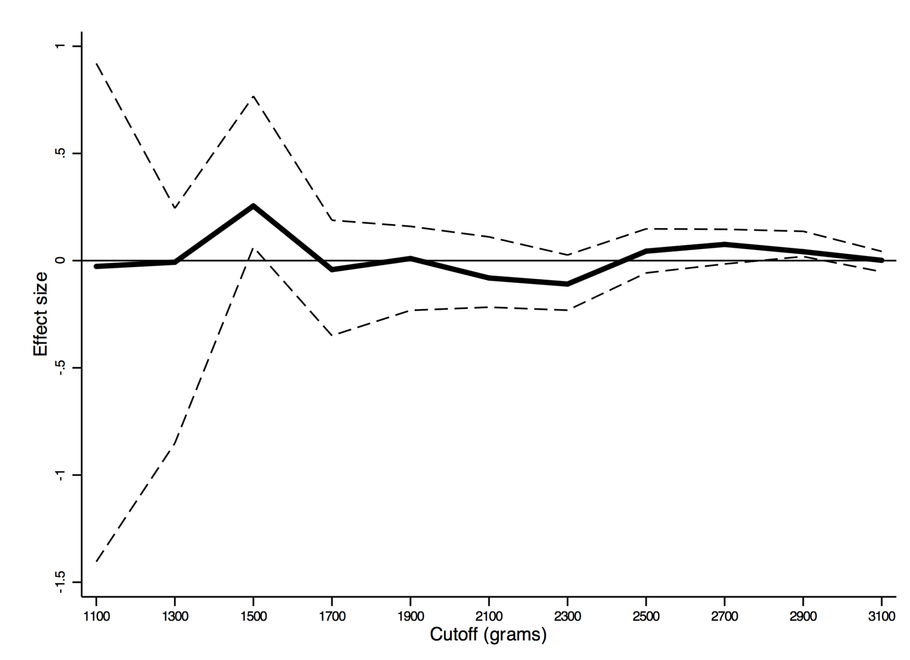
(d) Math test score, second degree polynomial

Notes: The solid lines plot conventional coefficient estimates of VLBW from local linear regressions similar to equation (1), using bandwidths between 100g and 300g in 10g intervals. The dotted lines plot the corresponding robust 95% confidence intervals. Panels (a) and (b), and panels (c) and (d) are drawn on the same scale, respectively.

Figure 4: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32+ sibling sample



(a) Language test score



(b) Math test score

Notes: The solid lines plot conventional coefficient estimates of VLBW from local linear regressions similar to equation (1), using a 200g bandwidth around cutoffs from 1,100g to 3,100g in increments of 200g. The dotted lines plot the corresponding robust 95% confidence intervals.

Figure 5: Discontinuities in sibling test scores at other points in the distribution of focal child birth weight, GA32+ sibling sample

Table 1: Distribution of covariates across the VLBW cutoff

	Gestational age of focal child: ≥ 32 weeks					Gestational age of focal child: < 32 weeks				
	Estimate	Bias-corrected estimate	Robust standard error	Mean of dependent variable	Obs.	Estimate	Bias-corrected estimate	Robust standard error	Mean of dependent variable	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Mother's characteristics										
Mother's age	1.118	[1.040]	(0.800)	27.735	2,156	0.285	[0.891]	(0.898)	27.942	1,520
Mother's education (years)	-0.246	[0.218]	(0.389)	11.239	2,156	0.121	[0.513]	(0.393)	11.275	1,520
Immigrant mother	-0.021**	[-0.052]	(0.027)	0.068	2,156	0.015	[0.020]	(0.044)	0.058	1,521
Married parents	0.047	[0.003]	(0.080)	0.535	2,156	-0.058	[-0.105]	(0.084)	0.513	1,521
B. Focal child characteristics										
Boy	-0.028	[-0.079]	(0.077)	0.456	2,156	-0.028	[-0.018]	(0.083)	0.599	1,521
Birth order	0.229	[0.166]	(0.173)	1.911	2,156	-0.002	[-0.060]	(0.169)	2.040	1,521
Multiple birth	0.065	[0.092]	(0.070)	0.208	2,156	-0.073	[-0.058]	(0.059)	0.151	1,521
C. Sibling characteristics										
Birth weight	-128.494*	[-188.938]	(105.751)	2898.702	3,210	-171.341	[-101.664]	(103.993)	3044.129	2,451
Boy	-0.003	[-0.033]	(0.068)	0.520	3,311	-0.006	[-0.053]	(0.066)	0.532	2,516
Multiple birth	0.026	[0.011]	(0.017)	0.023	3,311	0.031	[0.027]	(0.022)	0.018	2,516
Birth order	-0.115	[-0.154]	(0.147)	2.121	3,311	0.045	[0.005]	(0.139)	2.124	2,516
VLBW	0.012	[0.019]	(0.033)	0.046	3,311	0.018	[0.031]	(0.029)	0.041	2,516
Age difference - older sibling	-0.119	[-0.397]	(0.782)	6.586	1,634	0.803	[0.592]	(0.741)	6.056	1,382
Age difference - younger sibling	-0.400	[-0.691]	(0.449)	4.515	1,677	-0.752**	[-1.235]	(0.500)	4.417	1,134
D. Predicted sibling outcomes										
Language test score	-0.029	[0.015]	(0.049)	-0.120	3,210	-0.022	[0.045]	(0.047)	-0.123	2,449
Math test score	-0.044	[-0.018]	(0.056)	-0.165	3,210	-0.045	[0.010]	(0.049)	-0.152	2,449
Enrollment in academic track	-0.039	[-0.036]	(0.036)	0.387	3,210	-0.008	[0.030]	(0.034)	0.379	2,449
Enrollment in vocational track	0.016	[0.011]	(0.026)	0.438	3,210	-0.016	[-0.041]	(0.027)	0.439	2,449
Enrollment beyond compulsory schooling	-0.023	[-0.024]	(0.016)	0.771	3,210	-0.021	[-0.009]	(0.013)	0.764	2,449

Notes: Sample of (siblings of) focal children with birth weight within a 200g bandwidth around the 1,500g cutoff. Estimates in Panels A and B are based on the FC sample, while estimates in Panels C and D are based on the sibling sample (see Section 5). Columns 1-5 and Columns 6-10 in each row report results from separate local-linear regressions similar to equation (1) with outcome indicated in the row. All regressions use a triangular kernel and control for heaping at multiples of 50g. Columns 1 and 6 report conventional estimates, Columns 2 and 7 bias-corrected estimates, Columns 3 and 8 robust standard errors, Columns 4 and 9 the mean of the dependent variable for observations to the right of the cutoff, and Columns 5 and 10 the number of observations. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 2: Effects of early-life treatments on sibling academic achievement

	Gestational age of focal child	
	≥ 32 weeks	< 32 weeks
	(1)	(2)
Language test score	0.386**	-0.123
	[0.477]	[-0.042]
	(0.199)	(0.204)
Mean outcome	-0.155	-0.065
Observations	1,510	1,130
Math test score	0.255**	-0.085
	[0.415]	[-0.054]
	(0.180)	(0.189)
Mean outcome	-0.213	-0.117
Observations	1,516	1,139
Enrollment in academic track	0.098**	0.052
	[0.156]	[0.095]
	(0.074)	(0.070)
Mean outcome	0.432	0.446
Observations	2,759	2,120
Enrollment in vocational track	-0.034*	-0.056
	[-0.130]	[-0.086]
	(0.070)	(0.069)
Mean outcome	0.398	0.361
Observations	2,759	2,120
Enrollment beyond compulsory schooling	0.051	-0.027
	[0.017]	[0.007]
	(0.061)	(0.059)
Mean outcome	0.781	0.771
Observations	2,759	2,120

Notes: Sample of siblings of focal children with birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 3: Robustness of estimated spillover effects to model specification, siblings of GA32+ focal children

	Including controls	No heaping controls	Donut regressions	
			Excluding 1,500g	Excluding 1,490-1,510g
	(1)	(2)	(3)	(4)
Language test score	0.388***	0.350**	0.386**	0.380**
	[0.472]	[0.493]	[0.506]	[0.556]
	(0.182)	(0.199)	(0.239)	(0.269)
Mean outcome	-0.155	-0.155	-0.157	-0.155
Observations	1,510	1,510	1,442	1,407
Math test score	0.274***	0.310**	0.255**	0.297***
	[0.441]	[0.443]	[0.482]	[0.689]
	(0.157)	(0.180)	(0.201)	(0.243)
Mean outcome	-0.213	-0.213	-0.208	-0.209
Observations	1,516	1,516	1,448	1,413
	Rectangular kernel	CCT optimal bandwidth	Clustering	
			Birthweight	Mother
	(5)	(6)	(7)	(8)
Language test score	0.364**	0.291***	0.386***	0.386**
	[0.405]	[0.353]	[0.477]	[0.477]
	(0.177)	(0.128)	(0.131)	(0.214)
Mean outcome	-0.155	-0.153	-0.155	-0.155
Observations	1,510	2,414	1,510	1,510
Math test score	0.158*	0.211**	0.255*	0.255**
	[0.312]	[0.316]	[0.415]	[0.415]
	(0.162)	(0.130)	(0.233)	(0.199)
Mean outcome	-0.213	-0.207	-0.213	-0.213
Observations	1,516	1,887	1,516	1,516

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Additional controls included in column 1 are: focal child characteristics (gestational age and indicators for gender, birth order, multiple birth, year of birth, and region of birth), mother characteristics at the birth of the focal child (age, years of education, and indicators for immigrant status, marital status, and missing information on education), and sibling characteristics (birth weight and indicators for gender, birth order, multiple birth, and year of birth). Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 4: Robustness of estimated spillover effects to sample selection, siblings of GA32+ focal children

	Exclude VLBW siblings	Siblings of singleton focal children	Siblings of surviving focal children	Singleton siblings of surviving singleton focal children	Siblings of focal children with GA 32-33 weeks
	(1)	(2)	(3)	(4)	(5)
Language test score	0.401**	0.347**	0.403**	0.374**	0.505**
	[0.466]	[0.491]	[0.422]	[0.446]	[0.639]
	(0.205)	(0.209)	(0.210)	(0.226)	(0.294)
Mean outcome	-0.151	-0.155	-0.164	-0.155	-0.128
Observations	1,456	1,287	1,329	1,093	704
Math test score	0.263**	0.264**	0.299**	0.346**	0.265**
	[0.416]	[0.481]	[0.462]	[0.562]	[0.481]
	(0.183)	(0.201)	(0.198)	(0.231)	(0.243)
Mean outcome	-0.206	-0.205	-0.206	-0.184	-0.182
Observations	1,465	1,289	1,332	1,090	702

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 5: Discontinuities in health across the VLBW cutoff, siblings of GA32+ focal children

	(1)
Sibling admitted to hospital, focal child age 0-5	0.057
	[0.097]
	(0.066)
Mean outcome	0.374
Observations	3,311
Sibling admitted to hospital, focal child age 6-10	0.029
	[-0.015]
	(0.057)
Mean outcome	0.254
Observations	3,311
Sibling admitted to ER, focal child age 6-10	0.063*
	[0.183]
	(0.104)
Mean outcome	0.464
Observations	1,220

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 6: Family resources across the VLBW cutoff by age of focal child, GA32+ focal children

	Mother outcome		Father outcome		Family outcome	
	Age 0-5	Age 6-10	Age 0-5	Age 6-10	Age 0-5	Age 6-10
	(1)	(2)	(3)	(4)	(5)	(6)
Log income (thousands 2015 DKK)	0.041	0.247	-0.074	0.181*	-0.054	0.087
	[0.258]	[0.358]	[0.142]	[0.496]	[0.127]	[0.289]
	(0.257)	(0.285)	(0.266)	(0.297)	(0.193)	(0.228)
Mean outcome	4.494	4.462	5.386	5.103	5.949	5.795
Observations	2,152	2,122	2,109	2,071	2,154	2,142
Days worked per year	10.136*	4.674	5.910	13.461*		
	[23.706]	[8.061]	[19.628]	[29.794]		
	(13.866)	(14.942)	(14.808)	(15.577)		
Mean outcome	120.663	145.480	183.093	183.045		
Observations	2,151	2,119	2,108	2,070		
Labor force participation	-0.031	0.032	-0.024	0.052**		
	[-0.029]	[0.026]	[0.005]	[0.100]		
	(0.049)	(0.052)	(0.044)	(0.049)		
Mean outcome	0.874	0.841	0.914	0.868		
Observations	2,151	2,119	2,108	2,070		

Notes: Sample of focal children (with siblings) with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row averaged over the period indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for (parents of) focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 7: Family environment across the VLBW cutoff by age of focal child, GA32+ focal children

	Age 0-5	Age 6-10
	(1)	(2)
Divorce	0.098	0.010
	[0.081]	[0.018]
	(0.063)	(0.048)
Mean outcome	0.192	0.103
Observations	2,117	2,117
Mother's use of antidepressants	-0.060***	-0.033
	[-0.067]	[-0.044]
	(0.021)	(0.032)
Mean outcome	0.045	0.046
Observations	689	1,585
Father's use of antidepressants	0.006	0.032
	[0.023]	[0.051]
	(0.053)	(0.050)
Mean outcome	0.033	0.045
Observations	669	1,555

Notes: Sample of focal children (with siblings) with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Column 1 reports results over the age of 0-5 for divorce and 2-5 for antidepressant use. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row averaged over the period indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for (parents of) focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table 8: Heterogeneous spillover effects by sibship characteristics, siblings of GA32+ focal children

	Sibling gender vs focal child		Sibling gender	
	Different	Same	Female	Male
	(1)	(2)	(3)	(4)
Language test score	0.370*	0.403	0.280*	0.454
	[0.541]	[0.411]	[0.495]	[0.441]
	(0.281)	(0.289)	(0.271)	(0.292)
Mean outcome	-0.126	-0.184	0.017	-0.329
Observations	740	770	765	745
Math test score	0.075	0.434*	0.296***	0.208
	[0.400]	[0.435]	[0.696]	[0.163]
	(0.298)	(0.228)	(0.195)	(0.305)
Mean outcome	-0.198	-0.227	-0.289	-0.138
Observations	741	775	757	759

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Spillover Effects of Early-Life Medical Interventions

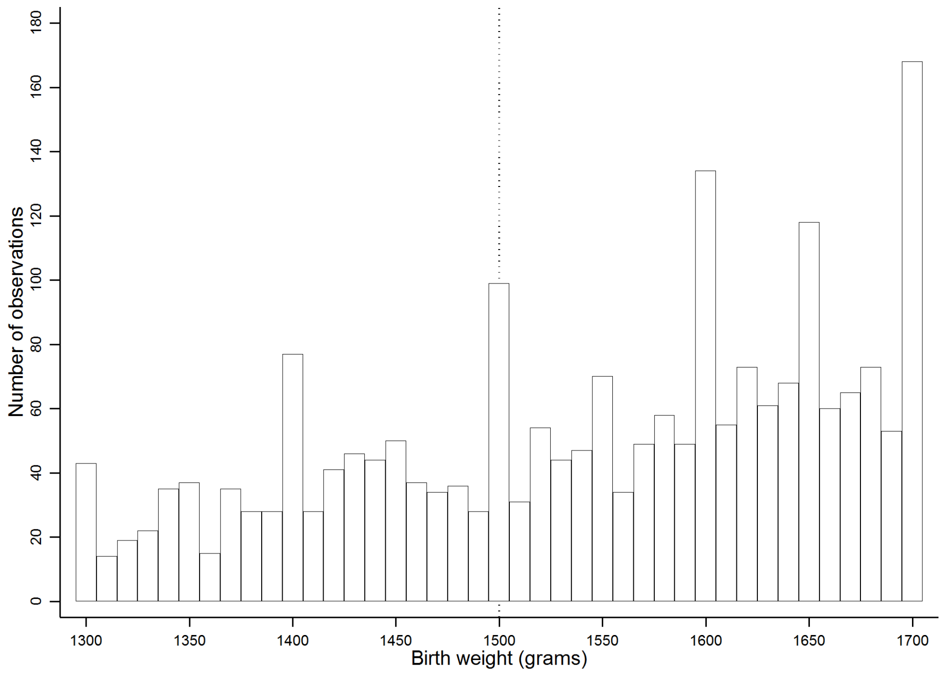
N. Meltem Daysal
University of Southern Denmark and IZA

Marianne Simonsen
Aarhus University and IZA

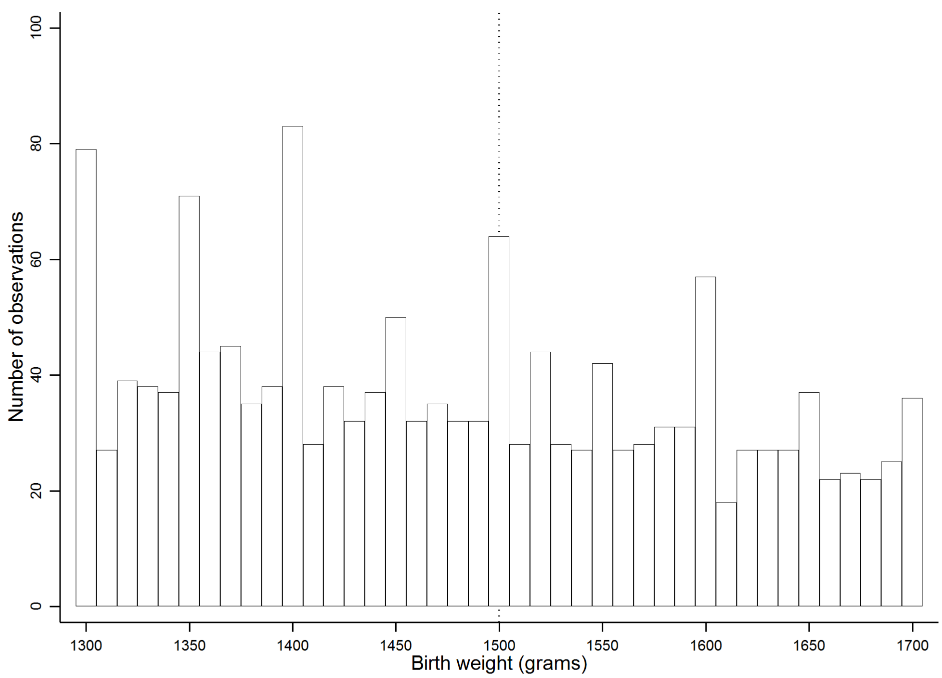
Mircea Trandafir
University of Southern Denmark and IZA

Sanni Breining
Ramboll Management Consulting

Online Appendix
(Not for publication)



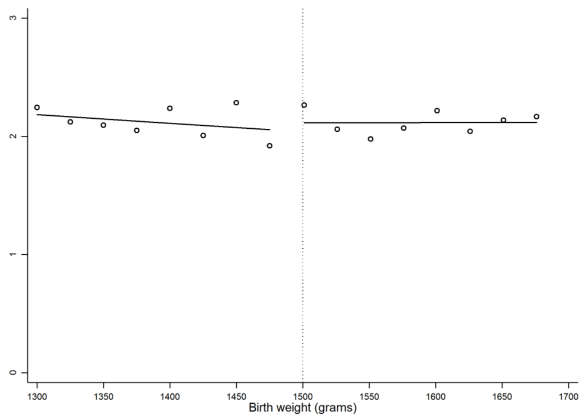
(a) Focal children with gestational age ≥ 32 weeks



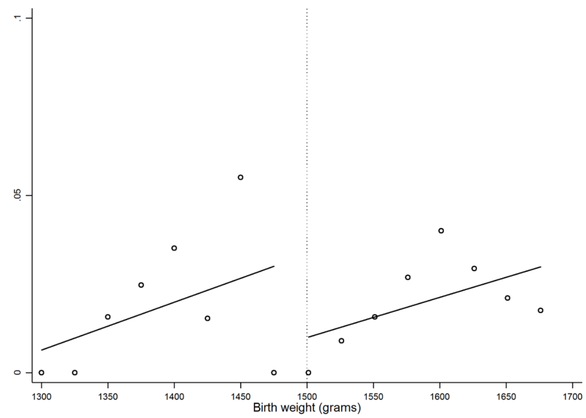
(b) Focal children with gestational age < 32 weeks

Notes: The corresponding estimates from local-linear regressions similar to equation (1) using the number of observations in a 10g bin as the dependent variable are: 0.092 (bias-corrected estimate -7.507, s.e. 6.955) and -1.806 (bias-corrected estimate -2.370, s.e. 3.965).

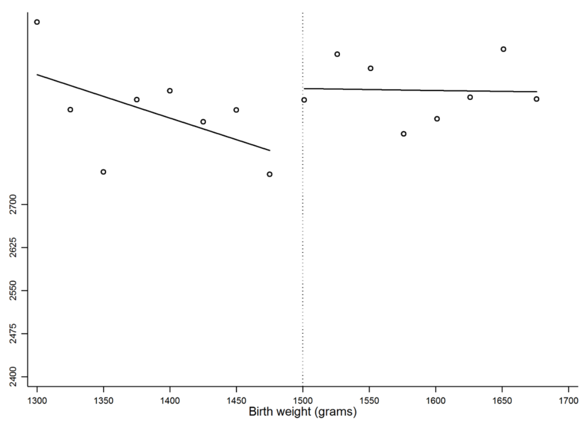
Figure A1: Frequency of observations around the VLBW cutoff, FC sample



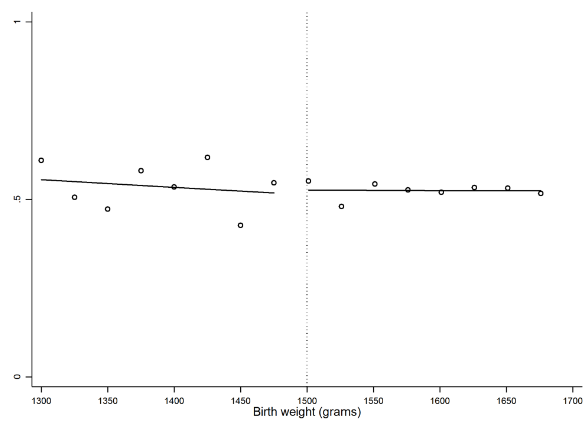
(a) Sibling parity



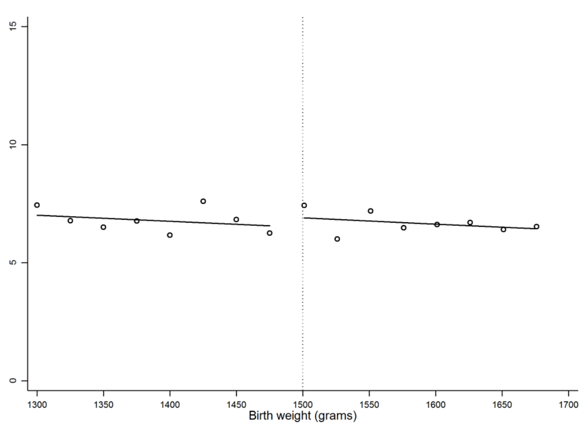
(b) Sibling plurality



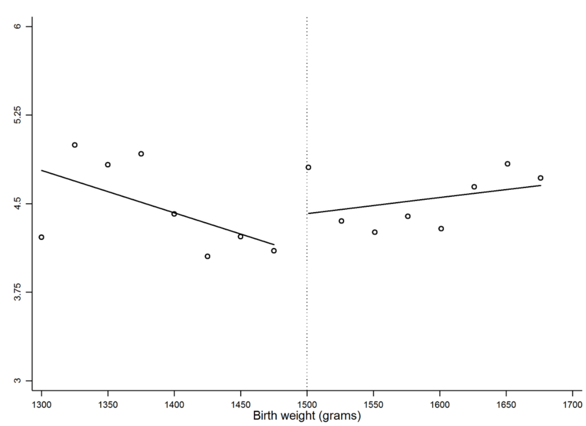
(c) Sibling birth weight



(d) Sibling male



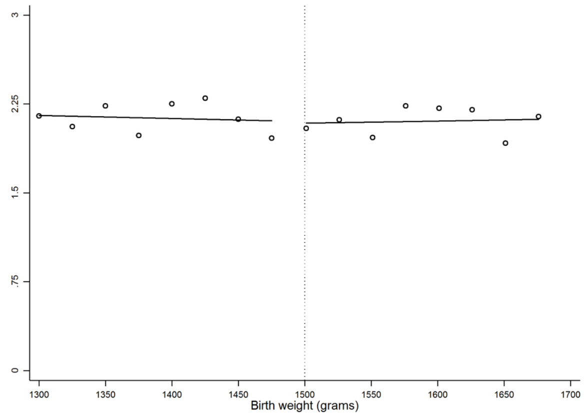
(e) Age difference, older siblings



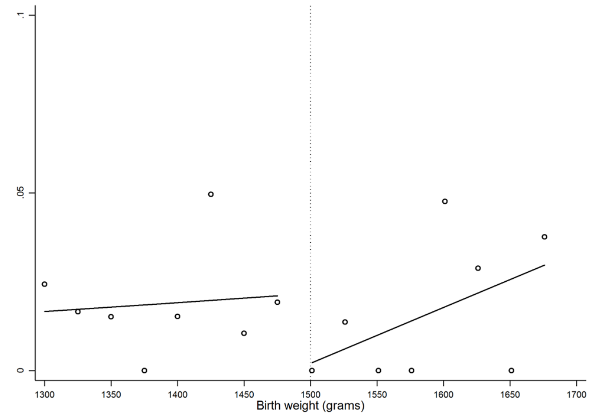
(f) Age difference, younger siblings

Notes: Each dot represents the average of the variable indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first degree polynomial estimated separately on either side of the VLBW cutoff.

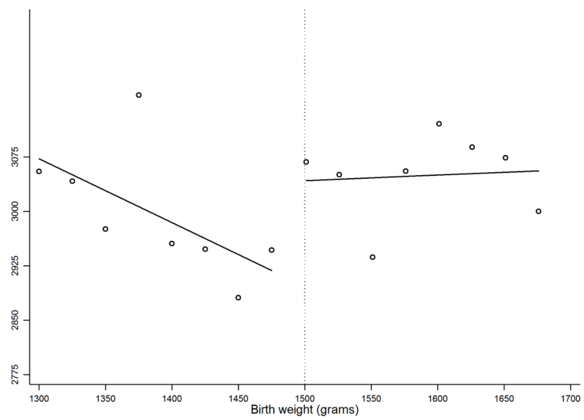
Figure A2: Distribution of selected covariates around VLBW cutoff, siblings of focal children with gestational age ≥ 32 weeks



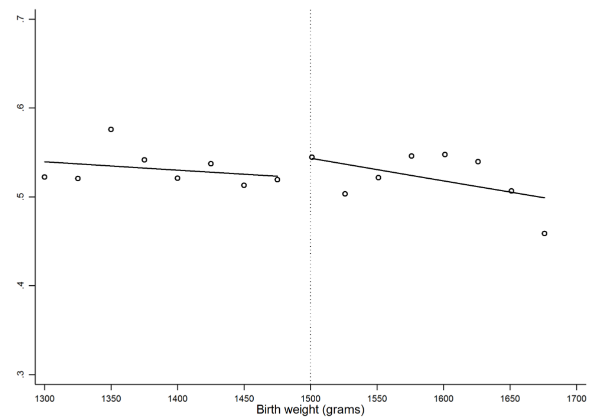
(a) Sibling parity



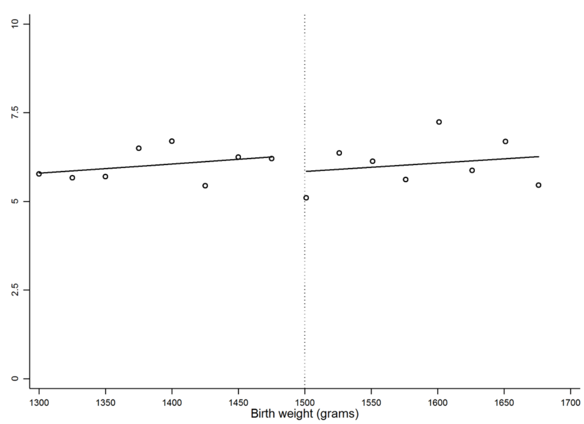
(b) Sibling plurality



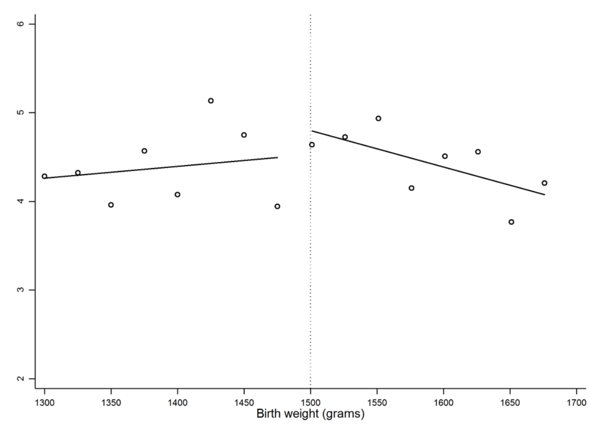
(c) Sibling birth weight



(d) Sibling male



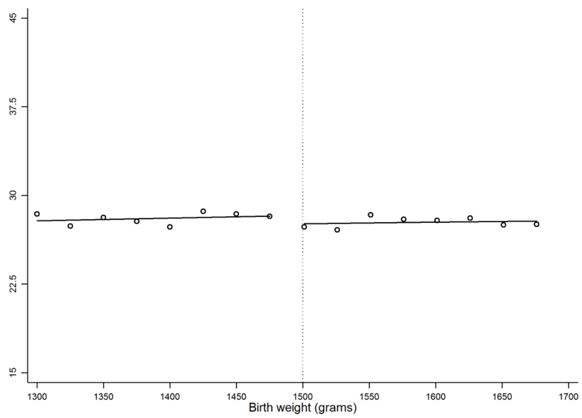
(e) Age difference, older siblings



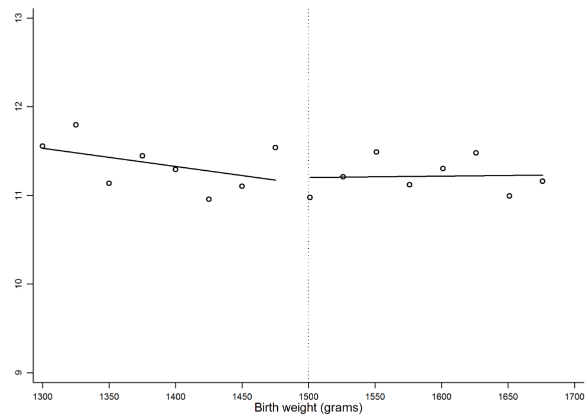
(f) Age difference, younger siblings

Notes: Each dot represents the average of the variable indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first degree polynomial estimated separately on either side of the VLBW cutoff.

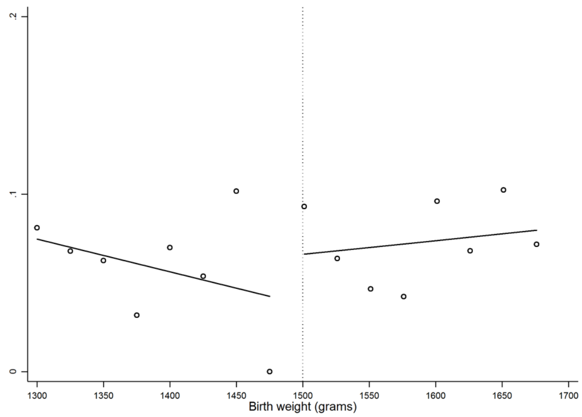
Figure A3: Distribution of selected covariates around VLBW cutoff, siblings of focal children with gestational age < 32 weeks



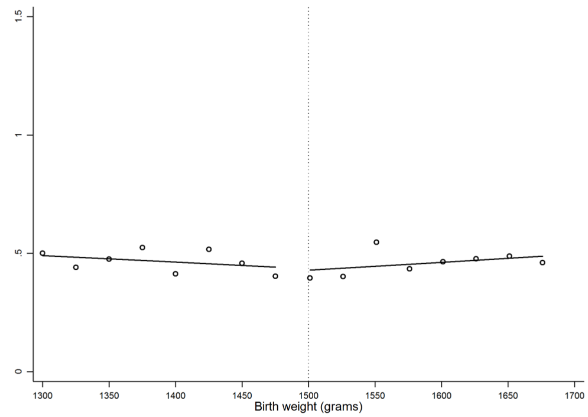
(a) Mother’s age at birth of focal child



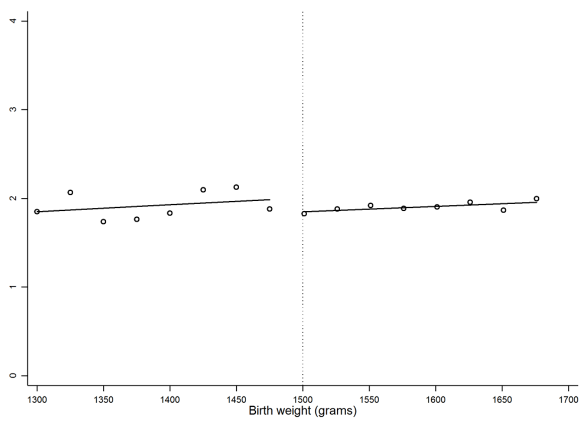
(b) Mother’s years of education



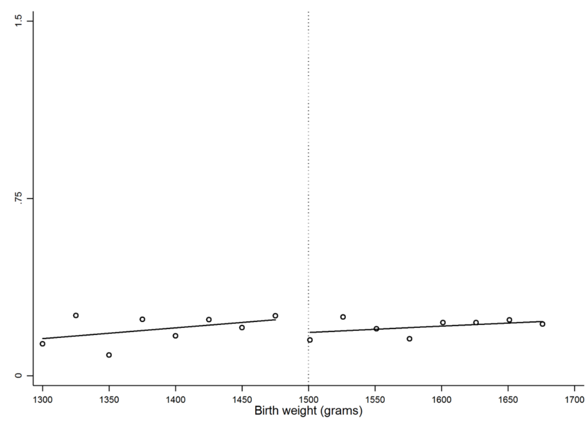
(c) Mother’s immigrant status



(d) Focal child male



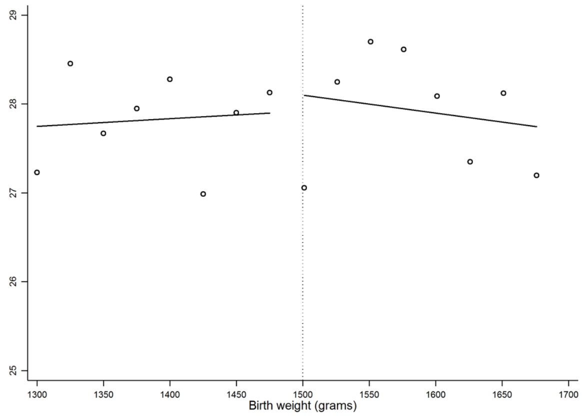
(e) Focal child parity



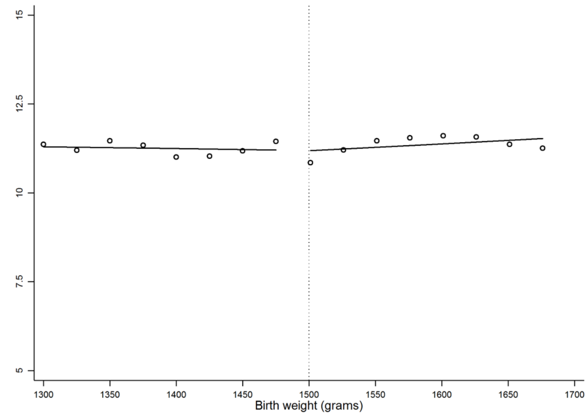
(f) Focal child plurality

Notes: Each dot represents the average of the variable indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first degree polynomial estimated separately on either side of the VLBW cutoff.

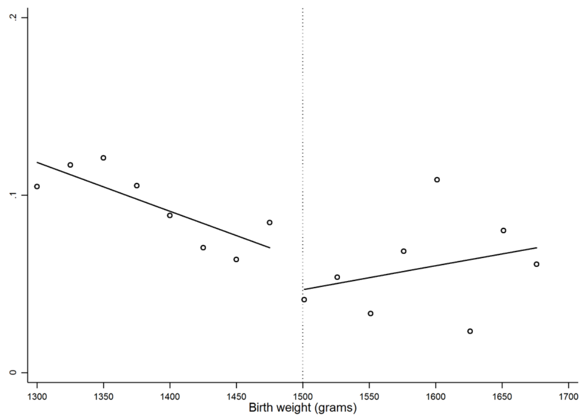
Figure A4: Distribution of selected covariates around VLBW cutoff, FC sample, children with gestational age ≥ 32 weeks



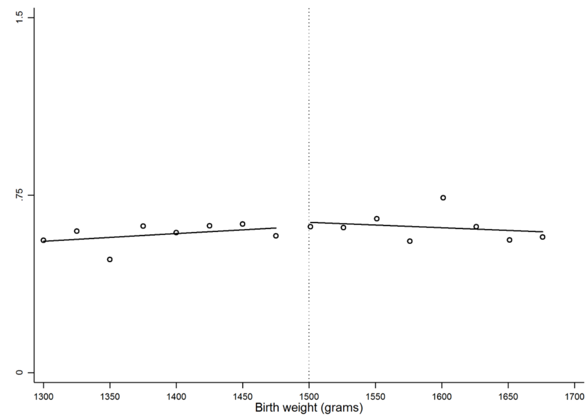
(a) Mother’s age at birth of focal child



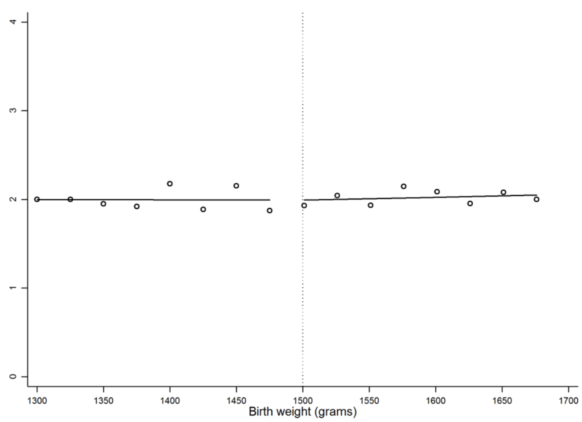
(b) Mother’s years of education



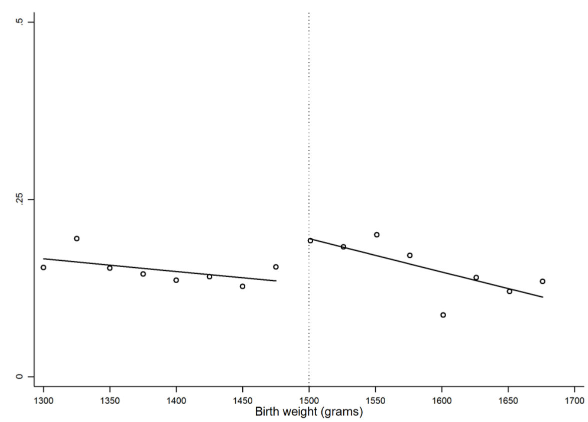
(c) Mother’s immigrant status



(d) Focal child male



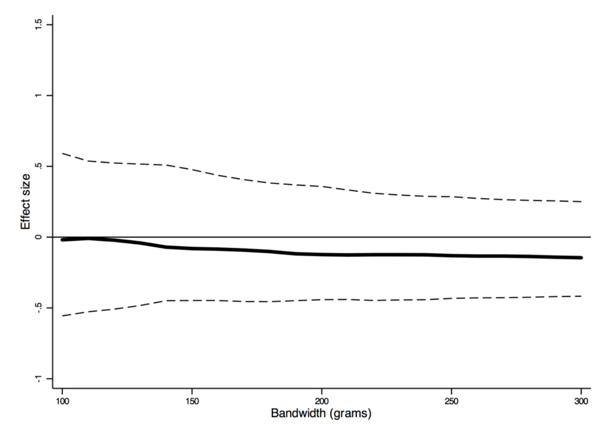
(e) Focal child parity



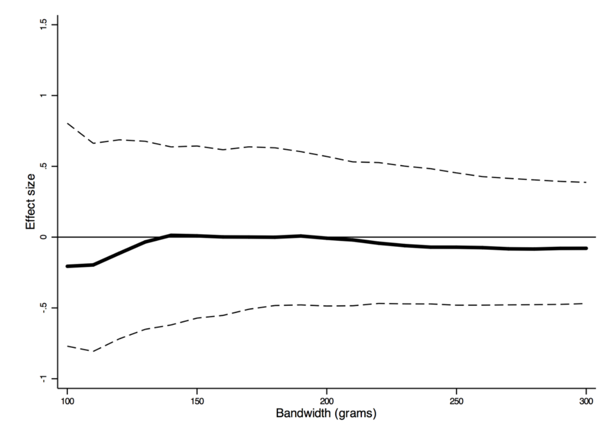
(f) Focal child plurality

Notes: Each dot represents the average of the variable indicated in the panel for a 40g bin. Siblings of focal children with birth weight of 1,500g are excluded. The lines plot a first degree polynomial estimated separately on either side of the VLBW cutoff.

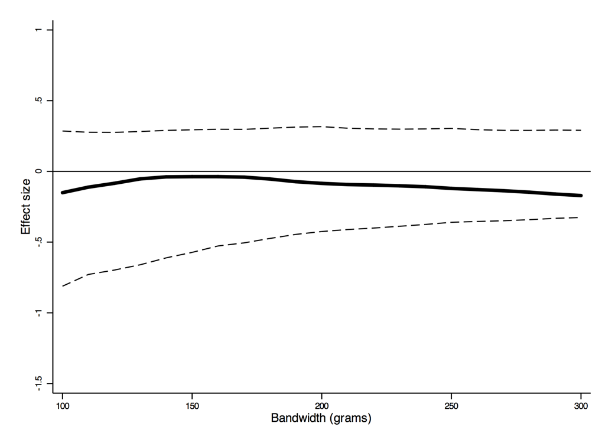
Figure A5: Distribution of selected covariates around VLBW cutoff, FC sample, children with gestational age $<$ 32 weeks



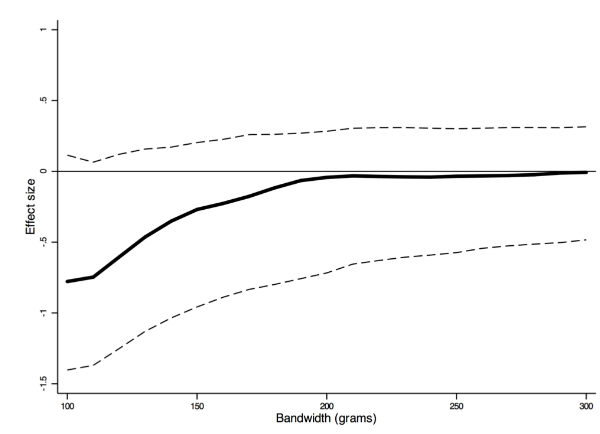
(a) Language test score, first degree polynomial



(b) Language test score, second degree polynomial



(c) Math test score, first degree polynomial



(d) Math test score, second degree polynomial

Notes: The solid lines plot conventional coefficient estimates of VLBW from local linear regressions similar to equation (1), using bandwidths between 100g and 300g in 10g intervals. The dotted line plot the corresponding robust 95% confidence intervals. Panels (a) and (b), and panels (c) and (d) are drawn on the same scale, respectively.

Figure A6: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32- sibling sample

Appendix Table A1: Sample construction

	Observations
Focal children	
Initial sample	772,998
Children missing information on birth weight or gestational age	73,385
Children with birth weight outside our bandwidth (1,300-1,700 grams)	695,014
Children with no siblings born within our sample period	922
Final sample of focal children with siblings	3,677
– with gestational age below 32 weeks	1,521
– with gestational age of at least 32 weeks	2,156
Siblings	
Initial sample (siblings of focal children above)	6,389
Siblings born after 1997 (no information on educational outcomes)	562
Final sample of siblings	5,827
– siblings of focal children with gestational age below 32 weeks	2,516
– siblings of focal children with gestational age of at least 32 weeks	3,311

Table A2: Distribution of covariates across the VLBW cutoff

	Gestational age of focal child: ≥ 32 weeks					Gestational age of focal child: < 32 weeks				
	Estimate	Bias-corrected estimate	Robust standard error	Mean of dependent variable	Obs.	Estimate	Bias-corrected estimate	Robust standard error	Mean of dependent variable	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Fertility										
Focal child birth order	0.229	[0.166]	(0.173)	1.911	2,156	-0.002	[-0.060]	(0.169)	2.040	1,521
Total family size	0.054	[0.009]	(0.160)	2.937	2,156	-0.075	[-0.158]	(0.163)	3.028	1,521
Number of younger siblings	-0.189	[-0.205]	(0.133)	0.899	2,156	0.046	[-0.006]	(0.147)	0.897	1,521
Probability of having younger siblings	-0.066	[-0.020]	(0.078)	0.611	2,156	0.077	[0.067]	(0.081)	0.602	1,521
B. Focal child test-taking behavior										
Age at test	-0.033	[-0.115]	(0.127)	16.137	1,004	0.083	[-0.011]	(0.116)	16.119	749
Probability of taking language test	0.014	[0.061]	(0.085)	0.682	1,400	0.149**	[0.190]	(0.087)	0.698	1,057
Probability of taking math test	-0.009	[0.022]	(0.089)	0.677	1,400	0.119*	[0.169]	(0.087)	0.706	1,057
C. Sibling test-taking behavior										
Age at test	-0.139	[-0.106]	(0.120)	16.035	1,602	0.001	[-0.066]	(0.106)	16.001	1,190
Probability of taking language test	0.029	[0.051]	(0.070)	0.808	1,877	0.009	[0.053]	(0.064)	0.787	1,447
Probability of taking math test	0.048	[0.069]	(0.068)	0.804	1,877	-0.003	[0.038]	(0.068)	0.792	1,447
D. Older sibling mortality										
28-day mortality	0.016	[0.027]	(0.020)	0.011	3,594	-0.005	[-0.007]	(0.016)	0.017	2,795
1-year mortality	0.021	[0.031]	(0.022)	0.015	3,594	-0.003	[-0.007]	(0.016)	0.022	2,795

Notes: Sample of (siblings of) focal children with birth weight within a 200g bandwidth around the 1,500g cutoff. Estimates in Panels A and B are based on the FC sample, estimates in Panels C are based on the sibling sample (see Section 5), and estimates in Panel D are based on a sample of all siblings belonging to the same cohorts as our sibling sample. Columns 1-5 and Columns 6-10 in each row report results from separate local-linear regressions similar to equation (1) with outcome indicated in the row. All regressions use a triangular kernel and control for heaping at multiples of 50g. Columns 1 and 6 report conventional estimates, Columns 2 and 7 bias-corrected estimates, Columns 3 and 8 robust standard errors, Columns 4 and 9 the mean of the dependent variable for observations to the right of the cutoff, and Columns 5 and 10 the number of observations. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table A3: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32+ sibling sample

	BW = 100		BW = 110		BW = 120		BW = 130		BW = 140	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language test score	0.502*	0.678	0.486*	0.686	0.470*	0.643	0.461**	0.596	0.455**	0.564*
	[0.556]	[0.576]	[0.525]	[0.649]	[0.529]	[0.594]	[0.528]	[0.583]	[0.521]	[0.583]
	(0.309)	(0.462)	(0.282)	(0.419)	(0.272)	(0.403)	(0.261)	(0.375)	(0.247)	(0.352)
Mean outcome	-0.122	-0.122	-0.116	-0.116	-0.106	-0.106	-0.098	-0.098	-0.115	-0.115
Observations	754	754	815	815	877	877	947	947	998	998
Math test score	0.444	0.038	0.452	0.106	0.449*	0.259	0.418*	0.401	0.395**	0.452
	[0.336]	[-0.245]	[0.388]	[-0.090]	[0.411]	[0.036]	[0.426]	[0.188]	[0.438]	[0.281]
	(0.267)	(0.357)	(0.246)	(0.336)	(0.239)	(0.329)	(0.230)	(0.313)	(0.220)	(0.298)
Mean outcome	-0.205	-0.205	-0.188	-0.188	-0.197	-0.197	-0.189	-0.189	-0.206	-0.206
Observations	758	758	818	818	881	881	953	953	1,004	1,004
	BW = 150		BW = 160		BW = 170		BW = 180		BW = 190	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
Language test score	0.437**	0.571*	0.425**	0.575*	0.413**	0.566*	0.402**	0.550*	0.390**	0.539*
	[0.522]	[0.575]	[0.515]	[0.565]	[0.497]	[0.589]	[0.493]	[0.576]	[0.488]	[0.568]
	(0.238)	(0.340)	(0.226)	(0.323)	(0.217)	(0.311)	(0.211)	(0.304)	(0.205)	(0.293)
Mean outcome	-0.144	-0.144	-0.151	-0.151	-0.141	-0.141	-0.144	-0.144	-0.136	-0.136
Observations	1,115	1,115	1,171	1,171	1,229	1,229	1,303	1,303	1,344	1,344
Math test score	0.366**	0.498	0.346**	0.524	0.317**	0.544	0.296**	0.532*	0.272**	0.522*
	[0.439]	[0.349]	[0.430]	[0.404]	[0.431]	[0.431]	[0.425]	[0.443]	[0.422]	[0.448]
	(0.212)	(0.290)	(0.202)	(0.278)	(0.194)	(0.270)	(0.189)	(0.265)	(0.184)	(0.257)
Mean outcome	-0.207	-0.207	-0.215	-0.215	-0.211	-0.211	-0.216	-0.216	-0.214	-0.214
Observations	1,119	1,119	1,175	1,175	1,233	1,233	1,307	1,307	1,348	1,348

Table A3: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32+ sibling sample (cont'd)

	BW = 200		BW = 210		BW = 220		BW = 230		BW = 240	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language test score	0.386**	0.518**	0.383**	0.505**	0.377**	0.498**	0.369**	0.492**	0.355**	0.495**
	[0.477]	[0.571]	[0.458]	[0.558]	[0.446]	[0.562]	[0.443]	[0.551]	[0.442]	[0.540]
	(0.199)	(0.284)	(0.190)	(0.269)	(0.185)	(0.264)	(0.182)	(0.259)	(0.178)	(0.251)
Mean outcome	-0.155	-0.155	-0.165	-0.165	-0.158	-0.158	-0.165	-0.165	-0.156	-0.156
Observations	1,510	1,510	1,561	1,561	1,626	1,626	1,705	1,705	1,766	1,766
Math test score	0.255**	0.507*	0.243**	0.496**	0.235**	0.472**	0.227**	0.446**	0.220**	0.418**
	[0.415]	[0.459]	[0.388]	[0.471]	[0.374]	[0.475]	[0.365]	[0.473]	[0.355]	[0.477]
	(0.180)	(0.251)	(0.172)	(0.238)	(0.168)	(0.234)	(0.164)	(0.230)	(0.161)	(0.224)
Mean outcome	-0.213	-0.213	-0.210	-0.210	-0.204	-0.204	-0.203	-0.203	-0.205	-0.205
Observations	1,516	1,516	1,566	1,566	1,630	1,630	1,706	1,706	1,767	1,767

	BW = 250		BW = 260		BW = 270		BW = 280		BW = 290	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
Language test score	0.341**	0.495**	0.328***	0.499**	0.319***	0.492**	0.312***	0.482**	0.303***	0.473**
	[0.447]	[0.520]	[0.448]	[0.505]	[0.444]	[0.503]	[0.436]	[0.511]	[0.431]	[0.509]
	(0.174)	(0.243)	(0.170)	(0.237)	(0.167)	(0.232)	(0.164)	(0.227)	(0.161)	(0.222)
Mean outcome	-0.156	-0.156	-0.155	-0.155	-0.159	-0.159	-0.155	-0.155	-0.154	-0.154
Observations	1,883	1,883	1,954	1,954	2,019	2,019	2,099	2,099	2,176	2,176
Math test score	0.211**	0.396**	0.203**	0.387**	0.194**	0.378**	0.186**	0.368**	0.175**	0.363**
	[0.346]	[0.478]	[0.336]	[0.468]	[0.329]	[0.459]	[0.322]	[0.455]	[0.321]	[0.437]
	(0.158)	(0.217)	(0.154)	(0.212)	(0.152)	(0.208)	(0.149)	(0.204)	(0.147)	(0.199)
Mean outcome	-0.204	-0.204	-0.207	-0.207	-0.211	-0.211	-0.207	-0.207	-0.210	-0.210
Observations	1,886	1,886	1,961	1,961	2,024	2,024	2,105	2,105	2,179	2,179

Table A3: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32+ sibling sample (cont'd)

	BW = 300	
	Poly1	Poly2
	(1)	(2)
Language test score	0.294***	0.470**
	[0.430]	[0.502]
	(0.158)	(0.218)
Mean outcome	-0.153	-0.153
Observations	2,413	2,413
Math test score	0.162**	0.365**
	[0.324]	[0.417]
	(0.144)	(0.196)
Mean outcome	-0.214	-0.214
Observations	2,416	2,416

Notes: Samples of siblings of focal children with gestational age of at least 32 weeks and birth weight within a bandwidth around the 1,500g cutoff indicated in the column. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row, with a polynomial in the running variable of order indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table A4: Discontinuities in sibling test scores at other points in the distribution of focal child birth weight, GA32+ sibling sample

	1,100g	1,300g	1,500g	1,700g	1,900g	2,100g	2,300g	2,500g	2,700g	2,900g	3,100g
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Language test score	0.080	0.182	0.386**	0.127	0.017	-0.094	-0.068	0.100**	0.039	-0.001	0.007
	[-0.080]	[0.042]	[0.477]	[0.177]	[0.006]	[-0.084]	[-0.089]	[0.115]	[0.045]	[0.020]	[0.010]
	(0.449)	(0.296)	(0.199)	(0.133)	(0.105)	(0.084)	(0.066)	(0.054)	(0.042)	(0.031)	(0.024)
Mean outcome	-0.160	-0.073	-0.155	-0.136	-0.134	-0.140	-0.143	-0.132	-0.123	-0.094	-0.064
Observations	380	789	1,510	2,767	4,876	8,321	14,845	27,308	50,138	86,921	129,755
Math test score	-0.027	-0.007	0.255**	-0.042	0.010	-0.081	-0.109	0.044	0.076	0.042***	0.001
	[-0.242]	[-0.304]	[0.415]	[-0.081]	[-0.036]	[-0.053]	[-0.102]	[0.045]	[0.065]	[0.078]	[-0.004]
	(0.593)	(0.280)	(0.180)	(0.137)	(0.100)	(0.084)	(0.066)	(0.052)	(0.041)	(0.030)	(0.024)
Mean outcome	-0.099	-0.131	-0.213	-0.171	-0.176	-0.211	-0.210	-0.204	-0.169	-0.128	-0.073
Observations	377	797	1,516	2,760	4,875	8,341	14,873	27,402	50,289	87,170	130,124

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the cutoff indicated in the column. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table A5: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32- sibling sample

	BW = 100		BW = 110		BW = 120		BW = 130		BW = 140	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language test score	-0.019	-0.206	-0.009	-0.196	-0.021	-0.114	-0.041	-0.034	-0.071	0.013
	[0.017]	[0.017]	[0.005]	[-0.072]	[0.007]	[-0.016]	[0.016]	[0.013]	[0.030]	[0.009]
	(0.293)	(0.402)	(0.272)	(0.375)	(0.263)	(0.359)	(0.255)	(0.339)	(0.244)	(0.321)
Mean outcome	-0.051	-0.051	-0.059	-0.059	-0.064	-0.064	-0.071	-0.071	-0.063	-0.063
Observations	577	577	619	619	670	670	724	724	773	773
Math test score	-0.150	-0.778*	-0.111	-0.747*	-0.084	-0.605	-0.053	-0.464	-0.039	-0.352
	[-0.263]	[-0.644]	[-0.226]	[-0.653]	[-0.211]	[-0.566]	[-0.189]	[-0.486]	[-0.161]	[-0.432]
	(0.280)	(0.387)	(0.256)	(0.366)	(0.248)	(0.350)	(0.240)	(0.329)	(0.230)	(0.308)
Mean outcome	-0.077	-0.077	-0.080	-0.080	-0.066	-0.066	-0.063	-0.063	-0.077	-0.077
Observations	575	575	617	617	669	669	725	725	777	777
	BW = 150		BW = 160		BW = 170		BW = 180		BW = 190	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
Language test score	-0.081	0.009	-0.085	0.001	-0.092	0.001	-0.102	-0.001	-0.118	0.008
	[0.015]	[0.036]	[-0.006]	[0.032]	[-0.024]	[0.064]	[-0.037]	[0.074]	[-0.040]	[0.062]
	(0.236)	(0.310)	(0.225)	(0.298)	(0.220)	(0.293)	(0.214)	(0.284)	(0.208)	(0.276)
Mean outcome	-0.057	-0.057	-0.068	-0.068	-0.078	-0.078	-0.092	-0.092	-0.066	-0.066
Observations	863	863	900	900	952	952	991	991	1,031	1,031
Math test score	-0.037	-0.270	-0.037	-0.227	-0.041	-0.177	-0.054	-0.116	-0.073	-0.065
	[-0.139]	[-0.377]	[-0.115]	[-0.332]	[-0.104]	[-0.288]	[-0.084]	[-0.269]	[-0.066]	[-0.244]
	(0.221)	(0.296)	(0.210)	(0.285)	(0.205)	(0.279)	(0.199)	(0.270)	(0.194)	(0.262)
Mean outcome	-0.085	-0.085	-0.092	-0.092	-0.102	-0.102	-0.115	-0.115	-0.106	-0.106
Observations	867	867	903	903	955	955	994	994	1,036	1,036

Table A5: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32- sibling sample (cont'd)

	BW = 200		BW = 210		BW = 220		BW = 230		BW = 240	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Language test score	-0.123	-0.008	-0.126	-0.020	-0.124	-0.043	-0.124	-0.060	-0.125	-0.071
	[-0.042]	[0.042]	[-0.054]	[0.023]	[-0.069]	[0.029]	[-0.073]	[0.015]	[-0.077]	[0.006]
	(0.204)	(0.269)	(0.197)	(0.259)	(0.193)	(0.254)	(0.189)	(0.248)	(0.186)	(0.244)
Mean outcome	-0.065	-0.065	-0.068	-0.068	-0.059	-0.059	-0.060	-0.060	-0.060	-0.060
Observations	1,130	1,130	1,174	1,174	1,238	1,238	1,272	1,272	1,317	1,317
Math test score	-0.085	-0.043	-0.093	-0.032	-0.097	-0.036	-0.102	-0.039	-0.109	-0.041
	[-0.054]	[-0.217]	[-0.053]	[-0.175]	[-0.050]	[-0.161]	[-0.045]	[-0.149]	[-0.037]	[-0.143]
	(0.189)	(0.255)	(0.183)	(0.245)	(0.179)	(0.239)	(0.175)	(0.234)	(0.172)	(0.229)
Mean outcome	-0.117	-0.117	-0.104	-0.104	-0.099	-0.099	-0.090	-0.090	-0.095	-0.095
Observations	1,139	1,139	1,180	1,180	1,244	1,244	1,280	1,280	1,326	1,326

	BW = 250		BW = 260		BW = 270		BW = 280		BW = 290	
	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2	Poly1	Poly2
Language test score	-0.131	-0.071	-0.134	-0.074	-0.134	-0.083	-0.137	-0.084	-0.142	-0.080
	[-0.074]	[-0.013]	[-0.078]	[-0.027]	[-0.082]	[-0.032]	[-0.083]	[-0.036]	[-0.082]	[-0.041]
	(0.183)	(0.238)	(0.179)	(0.232)	(0.177)	(0.228)	(0.175)	(0.225)	(0.172)	(0.222)
Mean outcome	-0.044	-0.044	-0.039	-0.039	-0.047	-0.047	-0.053	-0.053	-0.051	-0.051
Observations	1,413	1,413	1,447	1,447	1,473	1,473	1,504	1,504	1,532	1,532
Math test score	-0.120	-0.035	-0.129	-0.032	-0.137	-0.030	-0.147	-0.024	-0.160	-0.012
	[-0.028]	[-0.137]	[-0.030]	[-0.119]	[-0.029]	[-0.109]	[-0.026]	[-0.102]	[-0.020]	[-0.098]
	(0.169)	(0.223)	(0.166)	(0.216)	(0.163)	(0.213)	(0.161)	(0.210)	(0.159)	(0.207)
Mean outcome	-0.073	-0.073	-0.075	-0.075	-0.073	-0.073	-0.083	-0.083	-0.082	-0.082
Observations	1,423	1,423	1,457	1,457	1,484	1,484	1,515	1,515	1,543	1,543

Table A5: Robustness of estimated test score spillovers to the choice of bandwidth and degree of polynomial in focal child birth weight, GA32- sibling sample (cont'd)

	BW = 300	
	Poly1	Poly2
	(1)	(2)
Language test score	-0.145	-0.079
	[-0.083]	[-0.041]
	(0.170)	(0.218)
Mean outcome	-0.059	-0.059
Observations	1,630	1,630
Math test score	-0.171	-0.007
	[-0.018]	[-0.084]
	(0.157)	(0.204)
Mean outcome	-0.080	-0.080
Observations	1,640	1,640

Notes: Samples of siblings of focal children with gestational age of less than 32 weeks and birth weight within a bandwidth around the 1,500g cutoff indicated in the column. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row, with a polynomial in the running variable of order indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table A6: Robustness of estimated spillover effects to model specification, siblings of GA32- focal children

	Including controls	No heaping controls	Donut regressions	
			Excluding 1,500g	Excluding 1,490-1,510g
	(1)	(2)	(3)	(4)
Language test score	-0.183	-0.018	-0.123	-0.147
	[-0.195]	[0.093]	[-0.018]	[0.006]
	(0.191)	(0.204)	(0.248)	(0.305)
Mean outcome	-0.065	-0.065	-0.049	-0.054
Observations	1,130	1,130	1,089	1,039
Math test score	-0.039	-0.012	-0.085	0.019
	[-0.036]	[0.085]	[-0.046]	[0.220]
	(0.174)	(0.189)	(0.220)	(0.264)
Mean outcome	-0.117	-0.117	-0.109	-0.128
Observations	1,139	1,139	1,098	1,050
	Rectangular kernel	CCT optimal bandwidth	Clustering	
			Birthweight	Mother
	(5)	(6)	(7)	(8)
Language test score	-0.157	-0.148	-0.123	-0.123
	[-0.079]	[-0.180]	[-0.042]	[-0.042]
	(0.187)	(0.131)	(0.127)	(0.210)
Mean outcome	-0.065	-0.052	-0.065	-0.065
Observations	1,130	1,874	1,130	1,130
Math test score	-0.161	-0.198	-0.085	-0.085
	[-0.034]	[-0.189]	[-0.054]	[-0.054]
	(0.174)	(0.119)	(0.120)	(0.203)
Mean outcome	-0.117	-0.078	-0.117	-0.117
Observations	1,139	1,821	1,139	1,139

Notes: Sample of siblings of focal children with gestational age of less than 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Additional controls included in column 1 are: focal child characteristics (gestational age and indicators for gender, birth order, multiple birth, year of birth, and region of birth), mother characteristics at the birth of the focal child (age, years of education, and indicators for immigrant status, marital status, and missing information on education), and sibling characteristics (birth weight and indicators for gender, birth order, multiple birth, and year of birth). Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table A7: Robustness of estimated spillover effects to sample selection, siblings of GA32- focal children

	Exclude VLBW siblings	Siblings of singleton focal children	Siblings of surviving focal children	Singleton siblings of surviving singleton focal children	Siblings of focal children with GA 30-31 weeks
	(1)	(2)	(3)	(4)	(5)
Language test score	-0.124	-0.058	-0.174	-0.095	-0.130
	[-0.036]	[0.067]	[-0.073]	[0.047]	[-0.040]
	(0.210)	(0.211)	(0.223)	(0.230)	(0.228)
Mean outcome	-0.067	-0.086	-0.061	-0.074	-0.065
Observations	1,101	998	944	826	776
Math test score	-0.070	-0.081	-0.053	-0.066	-0.095
	[-0.022]	[0.005]	[-0.009]	[0.055]	[-0.126]
	(0.193)	(0.200)	(0.206)	(0.221)	(0.210)
Mean outcome	-0.113	-0.130	-0.124	-0.131	-0.120
Observations	1,108	1,007	948	829	788

Notes: Sample of siblings of focal children with gestational age of less than 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Appendix Table A8: Discontinuities in health and academic outcomes of focal children across the VLBW cutoff

	Gestational age	
	≥ 32 weeks	< 32 weeks
	(1)	(2)
A. Short-term health		
28-day mortality	-0.041**	-0.037
	[-0.077]	[-0.039]
	(0.031)	(0.045)
Mean outcome, non-VLBW focal children	0.062	0.072
Observations	2,156	1,521
1-year mortality	-0.054**	-0.019
	[-0.098]	[-0.004]
	(0.043)	(0.051)
Mean outcome, non-VLBW focal children	0.077	0.085
Observations	2,156	1,521
B. Academic achievement		
Language test score	0.229*	-0.136
	[0.419]	[-0.145]
	(0.235)	(0.245)
Mean outcome	-0.185	-0.044
Observations	939	697
Math test score	0.315**	-0.153
	[0.556]	[-0.261]
	(0.218)	(0.219)
Mean outcome	-0.259	-0.135
Observations	926	703

Appendix Table A8: Discontinuities in health and academic outcomes of focal children across the VLBW cutoff (cont'd)

	Gestational age	
	≥ 32 weeks	< 32 weeks
	(1)	(2)
C. Childhood disability		
ADHD diagnosis by age 10	0.004	0.018
	[-0.002]	[0.019]
	(0.015)	(0.013)
Mean outcome	0.008	0.010
Observations	2,156	1,521
Intellectual disability diagnosis by age 10	0.001	0.030
	[0.004]	[0.042]
	(0.019)	(0.029)
Mean outcome	0.015	0.007
Observations	2,156	1,521
Behavioral or emotional disorder diagnosis by age 10	0.008	0.033
	[0.006]	[0.039]
	(0.021)	(0.026)
Mean outcome	0.021	0.016
Observations	2,156	1,521
Epilepsy diagnosis by age 10	-0.001	0.037
	[-0.012]	[0.043]
	(0.012)	(0.033)
Mean outcome	0.018	0.030
Observations	2,156	1,521
Cerebral palsy diagnosis by age 10	0.021	0.055*
	[0.028]	[0.079]
	(0.020)	(0.044)
Mean outcome	0.017	0.052
Observations	2,156	1,521

Appendix Table A8: Discontinuities in health and academic outcomes of focal children across the VLBW cutoff (cont'd)

	Gestational age	
	≥ 32 weeks	< 32 weeks
	(1)	(2)
D. Health during school years		
Probability of a hospital admission between ages 6-15	-0.118**	-0.018
	[-0.163]	[0.016]
	(0.065)	(0.079)
Mean outcome	0.243	0.258
Observations	1,960	1,337
Probability of an ER visit between ages 6-15	-0.067	-0.006
	[-0.041]	[-0.039]
	(0.089)	(0.094)
Mean outcome	0.422	0.441
Observations	1,619	1,122

Notes: Sample of focal children (with siblings) with birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Table A9: Heterogeneous spillover effects by sibship characteristics, siblings of GA32+ focal children

	Birth order, sibling compared to focal child		Birth spacing	
	Younger	Older	≤ 3.5 years	> 3.5 years
	(1)	(2)	(3)	(4)
Language test score	0.417*	0.321***	0.145	0.557**
	[0.380]	[0.987]	[0.254]	[0.633]
	(0.229)	(0.378)	(0.268)	(0.319)
Mean outcome	-0.162	-0.122	-0.098	-0.197
Observations	1,265	245	648	862
Math test score	0.225	0.384*	0.160	0.265
	[0.313]	[0.882]	[0.283]	[0.450]
	(0.199)	(0.468)	(0.215)	(0.303)
Mean outcome	-0.234	-0.100	-0.149	-0.258
Observations	1,270	246	648	868

Notes: Sample of siblings of focal children with gestational age of at least 32 weeks and birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).

Appendix Table A10: Effects of early-life treatments on sibling academic achievement by sibling initial endowment

	Sibling birth weight	
	Above median	Below median
	(1)	(2)
Language test score	0.577*	0.302*
	[0.766]	[0.379]
	(0.443)	(0.228)
Mean outcome	-0.097	-0.175
Observations	401	1,109
Math test score	0.550**	0.142
	[0.901]	[0.272]
	(0.414)	(0.191)
Mean outcome	-0.013	-0.281
Observations	404	1,112

Notes: Sample of siblings of focal children with birth weight within a 200g bandwidth around the 1,500g cutoff. Each cell reports the estimated coefficient of the *VLBW* variable from a separate local-linear regression with a triangular kernel of the outcome listed in the row in the sample indicated in the column. Above and below median refer to the median birth weight in the sample of all children born within our study period, which is 3,480 grams. All regressions control for heaping at multiples of 50g. Bias-corrected estimates are listed in square brackets and robust standard errors in brackets below the coefficient estimates. Mean of the outcome is reported for siblings of focal children with birth weight above 1,500g. Stars indicate significance (*** significant at 1%, ** at 5%, * at 10%) based on robust confidence intervals centered on bias-corrected estimates (for details, see Calonico et al., 2014, 2018).